Modeling health and mortality dynamics, and their effects on public finance

Yang, Y.

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

YING YANG

Modeling Health and Mortality Dynamics, and Their Effects on Public Finance

Modeling Health and Mortality Dynamics, and Their Effects on Public Finance

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Ruth First zaal van de Universiteit op dinsdag 2 september 2014 om 10.15 uur door

YING YANG

geboren op 19 juli 1985 te Jiamusi City, China.

PROMOTIECOMMISSIE:

PROMOTORES:	prof. dr. Anja M.B. De Waegenaere
	prof. dr. Bertrand Melenberg
OVERIGE LEDEN:	dr. Katrien Antonio
	prof. dr. Peter Kooreman
	dr. Wilma J. Nusselder
	dr. Martin Salm

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude and thanks to my supervisors, Anja de Waegenaere and Bertrand Melenberg. They have been motivating and understanding mentors. I am very grateful for their continuous guidance and support during the past years. They helped me with my growth in research, in career, and in my life. They were always patient and generous with their time. The knowledge and research attitude I learned from them are priceless for me. I greatly appreciate their devotion in helping me to make this dissertation possible.

I would also like to thank my doctoral committee members, Katrien Antonio, Peter Kooreman, Wilma Nusselder, and Martin Salm for their valuable comments and suggestions.

I am very grateful to many friends who made me at ease while living in a foreign country. Their support and enthusiasm made Tilburg my second home. A special thanks to Rasa - officemate, friend, and "sister"- who always believed in me. Thank you for your encouragement, caring, listening, and being there for me. A special thanks to Kim, Lisanne, Jiehui, Christane, Yachang, Ivo, Edith, and Luc for all the pleasant time we have spent, for all the discussions about research, and for all the enjoyable chats. I also would like to thank Martin, Tobias, Otilia, and Meltem for their important advices during my job search. To my Chinese friends, Ruixin, Juanjuan, Huaxiang, Fangfang, Geng, Yan, and many others, thank you for the relaxing Chinese parties. I would also like to thank my other fellows, Gaia, Sara, Jan, Jaroslav, and Marco. I really enjoy our lunches and other social activities.

I thank our secretaries, Korine, Anja, Heidi, and Lenie for their friendly help whenever I needed, especially for their long distance help after I moved to the United States.

Particularly, I want to express my special appreciation to my parents in China for their understanding of me being so far away from them, for their unconditional support, and for all the sacrifices they made for me. Finally, I would like to thank Jia,

who always encourages and supports me in my struggle for finishing my dissertation during the past one year and a half.

Above all, I would like to thank all individuals who supported me in writing and encouraged me to strive towards my goal.

Ying Yang
Harbin, China, July 2014

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Many countries in the world have experienced increases in life expectancy and the accompanying population ageing over the past century. Saving sufficiently for retirement, being able to face higher pension expenses, efficiently allocating health care resources are significant challenges for individuals, public and private pension funds, insurance companies, and government (see, e.g., Bloom, Canning, Mansfield, and Moore, 2007; Hári, De Waegenaere, Melenberg, and Nijman, 2008; Pitacco, Denuit, Haberman, and Olivieri, 2009b). Like many other countries, the United States is facing a shift in the demographic structure of the population. The percentage of the population aged 65 and over has increased from 9.2% in 1960[1] to 13.1% in 2010[2]. To reduce the accompanying increased public pension expenditure, the U.S. 1983 Social Security Amendments has raised the full retirement age for cohorts born after 1937 gradually from 65 years in 2002 to 67 years in 2026. A concern associated with such policies is that even if an increase in retirement age effectively relieves the increased pension liability, possible spillover effects, such as increases in spending of disability insurance and social security insurance, and health expenditure may offset the reduced pension expenditure if people are not healthy enough to work (see,e.g., Munnell, Meme, Jivan, and Cahill, 2004; Munnell and Libby, 2007; Cutler, Meara, and Richards-Shubik, 2011; Unger and Schulze, 2013). This means that the complexity of changes in public finance associ-

[1]Population estimates provided by the U.S. Census Bureau. See `http://www.census.gov/popest/data/national/asrh/pre-1980/PE-11.html`.

[2]Data Source: A Profile of older American: 2011, Department of Health & Human Services. See `http://www.aoa.gov/Aging_Statistics/Profile/2011/4.aspx`

ated with ageing is not only caused by people's longer lifetimes, but also by the future development of people's health. Therefore, an effective and efficient policy making process requires not only quantifying the increase in life expectancy, which measures the expected remaining years of life at a given age and time, but also considering the development of healthy life expectancy, which measures the expected remaining lifetime in good health.

Another important concern relating to a large part of government expenditure in the United States is the fast growing healthcare expenditure over the past 50 years. The healthcare spending in the United States as a share of GDP (gross domestic product) has increased from 5.2% in 1960 to 17.9% in 2011[3]. The rising healthcare cost continuously takes up a larger proportion of the annual government budget, and aggravates the government burden considerably. There is a growing stream of literature that investigates whether, and to what extent, various factors determine the growth of healthcare expenditure. For instance, Hansen and King (1996), Manton, Lamb, and Gu (2007), Moscone and Tosetti (2010), Xu, Saksena, and Holly (2011), Solakoglu and Civan (2012), and many others suggest that national income, the price of healthcare, public financing, age structure, and population health are important factors affecting the growth of health expenditure. As one might assume, the use of healthcare services depends on people's health condition. Solakoglu and Civan (2012) adopt population health as an indicator of healthcare need, and find that the rising share of healthcare expenditure in GDP can be explained by the growing healthcare need. Therefore, better understanding of the development of population health provides relevant information for policy makers to improve decisions when allocating scarce healthcare resources.

In light of these concerns, the primary motivation of this dissertation is to provide insights into the future developments of mortality and population health, and the associated effects on public finance in the United States. Chapter 2 models the future developments of population health and quantifies the degree of uncertainty in the future developments. Chapter 3 jointly models the future developments of mortality and health, using a similar approach as for health in Chapter 2. This allows us to further investigate the association between the developments of life expectancy and healthy life expectancy, taking into account dependence between developments of mortality and health. Chapter 4 extends the forecast model developed in Chapter 3 by taking into account the dependence between male and female mortality and health. The model is used to estimate the effects on (healthy) life expectancy of a policy that links the retirement age to life expectancy. Finally, Chapter 5 studies another important part of public finance, the growth of healthcare expenditure. In this chapter, we investigate the dynamic relationship between the growth of healthcare cost and a relatively large set of its determinants, with special attention on the effect of people's health on the growing

[3]Data Source: "NHE summary including share of GDP, CY 1960-2011" provided by National Health Expenditure Data from Center of Medicare and Medicaid.

healthcare cost. The rest of the introduction will address each chapter in detail.

There is an extensive literature on modeling trends in population health. However, forecasting future health developments is not a trivial task. Several factors may affect population health in different directions and complicate the prediction of health changes. For example, Ruhm (2000) and Erdil and Yetkiner (2009) suggest that economic growth has a positive effect on the people's health. Michaud, Goldman, Lakdawalla, Zheng, and Gailey (2009) find that, on the one hand, increased obesity reduces life expectancy and increases morbidity; on the other hand, reduced smoking shows an opposite effect. The net effect remains unclear, and is surrounded by a lot of uncertainty. Much of the literature on forecasting health uses a deterministic approach (see, e.g., Singer and Manton, 1998; Jacobzone, 2000; Jagger, Matthews, Spiers, Brayne, Comas-Herrera, Robinson, Lindesay, and Croft, 2006; Manton, Gu, and Lamb, 2006a; Manton, Lamb, and Gu, 2007). Commonly used deterministic approaches are to assume population health improves with a certain speed annually, or to consider a number of deterministic scenarios for the development of health. One shortcoming of such deterministic approaches is that they do not provide information regarding the likelihood of changes in population health. Exceptions are Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012) and van Baal, Peters, Mackenbach, and Nusselder (2013). These studies develop Lee and Carter (1992) type approaches to model health transition probabilities and disability rates for the Dutch population. A major advantage of a Lee and Carter type approach is that it provides not only the forecasts of the future health changes, but also the corresponding uncertainties. Chapter 2 of this dissertation, which is based on the working paper, Yang, De Waegenaere, and Melenberg (2013b), extends the Lee and Carter (1992) approach by including observed variables, namely GDP per capita and the unemployment rate to model and forecast the health changes of the U.S. population. An important advantage of including observed variables is that future forecasts not only depend on estimated latent time trend but also on the (future) developments of GDP and unemployment rate. Moreover, because of taking into account additional information besides the latent time trend as in the original Lee-Carter model, this model might generate more precise model-based forecasts.

In Chapter 3, the future development of (healthy) life expectancy is examined, distinguished by genders. This chapter is based on the working paper, Yang, De Waegenaere, and Melenberg (2013a). Healthy life expectancy now is often used to measure the quality of life (van de Water, Perenboom, and Boshuizen (1996)) and measures changes in population health (Laditka and Laditka (2002)). In many studies, forecasts of healthy (or disability free) life expectancy allow changes in future mortality rates, but assume that future health remains constant or has multiple deterministic scenarios (see for example, Jagger, Matthews, Spiers, Brayne, Comas-Herrera, Robinson, Lindesay, and Croft (2006), Jacobzone (2000), and Manton, Gu, and Lamb (2006a)). As a consequence, the uncertainty of future healthy life expectancy generated by the uncer-

tainty of future health changes, which is possibly larger than the uncertainty of future mortality, cannot be sufficiently quantified. Moreover, separately treating mortality and health may result in biased estimation of life expectancy due to the possible high dependence of mortality on health. Applying the methodology proposed in Chapter 2, and extending it to jointly model mortality and health, Chapter 3 forecasts the development of future (healthy) life expectancy by taking into account the joint dynamics of mortality, health, and macroeconomic variables, quantifying its future uncertainties derived from both mortality and health. Moreover, it is well-documented that patterns of mortality and health are not the same for males and females (see, for example, Van Oyen, Cox, Jagger, Cambois, Nusselder, Gilles, and Robine (2010) and Van Oyen, Nusselder, Jagger, Kolip, Cambois, and Robine (2013)). Therefore, we also briefly discuss the gender disparities in (healthy) life expectancy in Chapter 3.

As discussed earlier in this introduction, one possible policy to deal with consequences of continuing increases in life expectancy is to raise the retirement age. Including the United States, several countries have already implemented such policies (see OECD (2013) and Schwan and Sail (2013)). One drawback of deterministically raising the retirement age for some period in advance is that there is a considerable degree of uncertainty regarding the development of future life expectancy. For this reason, countries such as Italy, Denmark, Greece, the Netherlands, Slovakia, and Cyprus have implemented policies in which the development of retirement age is directly linked to the development of life expectancy (see, e.g., Schwan and Sail, 2013). However, whether trends in health of the elderly support the raise in retirement age is a question that is not addressed in many studies. In addition to keeping people's remaining lifetime after retirement stable over time in order to avoid further increases in pension liabilities, two aspects related to retirees' health deserve attention. First, the number of years that retirees can enjoy retirement in good health might be significantly impacted by the policy if healthy life expectancy does not grow at the same pace as life expectancy (see, e.g., Cutler, Meara, and Richards-Shubik, 2011). Second, if the fraction of individuals that is sufficiently healthy to work until retirement would decrease significantly due to the policy, increases in unemployment or other social security benefits might offset the benefits of the policy in terms of reduced pension payments (see, e.g., Munnell, Meme, Jivan, and Cahill, 2004). Whether people will be healthy enough to work longer is a concern for the policy makers (see, e.g., Munnell, Meme, Jivan, and Cahill, 2004; Munnell and Libby, 2007; Unger and Schulze, 2013). Therefore, Chapter 4 of this dissertation, which is based on the working paper De Waegenaere, Melenberg, and Yang (2014), estimates the effects of a retirement age policy in which, roughly speaking, an increase in life expectancy of one month is accompanied by an increase in retirement age of one month. We investigate the effects of such a policy on life expectancy and healthy life expectancy before and after retirement, as well as the likelihood of being in good health at retirement age. To do so, the model developed in Chapter 3 is extended

to jointly model trends in mortality and health of both genders.

In Chapter 5, which is based on the working paper, Yang and Melenberg (2014), we investigate the development of the U.S. healthcare costs. The U.S. healthcare costs represent a significant part of the country's GDP, it is 17.9% in 2011[4]. As suggested by the literature, important factors affecting healthcare costs include national income (Christiansen, Bech, and Lauridsen (2007) and Amiri and Ventelou (2012)), demographic structure (Xu, Saksena, and Holly (2011)), healthcare price (Murthy and Ukpolo (1994)), public financing (Gerdtham and Jonsson (2000) and Murthy and Okunade (2000)), and technological progress (Berndt, Cutler, Frank, Griliches, Newhouse, and Triplett (2000) and van Elk, Mot, and Franses (2009)). Moreover, the ageing of the population and people's health play a major role in the future development of healthcare costs (See Solakoglu and Civan (2012), Dreger and Reimers (2005), and Murthy and Okunade (2000)). As the demand for healthcare is ultimately derived from the demand for better health, we examine in this chapter the health status of the elderly together with macroeconomic determinants and the age structure of the population as drivers of the healthcare spending growth. There are several complications to be dealt with when analyzing the relationship between health expenditure and its determinants. First, most of the studies only include a few factors, omitting important determinants, possibly resulting in an omitted variable bias when quantifying, for example, the income elasticity (Roberts (1999) and Gerdtham and Jonsson (2000)), or the effect of population ageing (Zweifel, Felder, and Meiers (1999) and Yang, Norton, and Stearns (2003)). Second, there may exist simultaneous relationships between healthcare spending and its determinants. For example, possibly a bilateral relationship exists between health expenditure and the elderly's health condition. On the one hand, an increase in the population fraction of the elderly in good health may reduce the need for healthcare services, which in turn might slow down the growth in healthcare cost; on the other hand, part of the increased healthcare expenditure may be attributed to better medical treatments and provisions of services to maintain life quality, which may improve people's health. Moreover, a reverse effect may exist from increased health expenditure on the growth of national income, through the enhancement of education, improvement in labor participation, and higher productivity due to health improvement brought by higher health expenditure (Erdil and Yetkiner (2009)). As a result, failure to take into account possible simultaneous relationships may under- or overestimate the effects of the variables of interest. Finally, an application of the appropriate methodology in this study turns out to be challenging. Trends in health expenditure and its determinants indicate non-stationarity. We find different forms of nonstationarity. Such different forms of nonstationarity complicate the econometric analysis considerably. The literature also documents conflicting conclusions regarding the stationarity/non-

[4]Data source: "NHE summary including share of GDP, CY 1960-2011" provided by National Health Expenditure Data from Center of Medicare and Medicaid.

stationarity of health expenditure and its determinants, and possible cointegration relationships between them (see reviews provided by Gerdtham and Lothgren (2000) and Moscone and Tosetti (2010)). Chapter 5 proceeds as follows. First, a relatively large number of factors are considered when investigating the changes in health expenditure. Second, we apply variable-specific transformations so that after applying these transformations the transformed variables are (close to) stationary. Third, we apply a Vector Auto Regression (VAR) model to capture the joint dynamic relationships between these stationarized variables. In the VAR framework, we allow the other variables to influence health expenditure and also the other way around. In this way, we try to reduce potential biases due to omitted variables, and avoid ignoring mutual relationships. We further evaluate forecasts of the healthcare spending based on an out-of-sample analysis. In particular, we compare the forecasts derived from the VAR model including the elderly's health status, with forecasts ignoring the elderly's health status, and with "official forecasts" published by Center of Medicare and Medicaid (CMS). We find that forecasts in the VAR model by taking into account the elderly's health status are superior than forecasts from the other two compared methods.

CHAPTER 2

STOCHASTIC MODELING AND FORECASTING OF HEALTH CHANGES IN THE U.S. POPULATION

This Chapter is based on Yang, De Waegenaere, and Melenberg (2013b)

This chapter proposes a model for self-assessed health at an aggregate level that allows to generate age- and gender-specific stochastic forecasts of future health. We decompose health status into a time effect and an age effect. We then further decompose the time effect into observed macroeconomic quantities (GDP and unemployment rate) and an unobserved latent time factor. We use data on the U.S. population's self-assessed health for both males and females to estimate the model. The estimation results show that trends in health can be largely captured by trends in the observed macroeconomic quantities. Next, based on forecasts of the observed and the unobserved time effects, using a vector auto regression (VAR) model, we present forecasts for future health together with the corresponding forecasting uncertainty, showing that there is no clear future trend upward or downward. A backtesting analysis suggests that our approach with macroeconomic quantities significantly improves the forecasting accuracy for future health development compared with a simple extrapolation based approach. It also outperforms the model without taking into account observed variables.

2.1 Introduction

Over the past century, understanding and predicting health changes in the United States has gained growing interest, not only from demographers and health economists,

but also from institutions, such as insurance companies, pension funds, social security, and government. For example, the United States' total spending for health care as a share of GDP (gross domestic product) is the highest among the OECD (Organisation for Economic Co-operation and Development) countries. It is almost double the OECD average and still growing. As health care expenditure generally increases with age and bad health status, it is important for institutions, such as health service providers, to assess to what extent health will change in the future. Moreover, better understanding health changes might be helpful for policy makers to improve labor participation decisions. For instance, many countries currently start increasing the retirement age gradually in order to reduce the rising pension costs because of an increase in life expectancy. However, such a policy decision might be inconsiderate if it only relies on the information of life expectancy, and ignoring people's future health changes. Since a rise in the retirement age may have an adverse effect if people are not healthy enough to work longer, it may lead to higher government spending on healthcare or disability benefits, possibly offsetting the reduced pension costs. A basic ingredient here is a good understanding of the development in health, now and in the future.

Future health changes in the U.S., however, are not trivial to predict. Costa (2002) states that functional limitations of the older U.S. men has reduced annually from the early twentieth century to the early 1990s. Moreover, the health of adults aged 50-64 has improved on average from 1984 to 2001, examined by Duggan and Imberman (2006) using self-assessed health from the National Health Interview Survey (NHIS). However, the future development of health is quite uncertain. Health might be affected by many factors, such as the economic situation, technological advances, strengthening of primary healthcare, and people's lifestyle choices. These factors may have large and offsetting effects. For instance, Michaud, Goldman, Lakdawalla, Zheng, and Gailey (2009) find that, on the one hand, increased obesity reduces life expectancy and increases morbidity for a number of years before death and, on the other hand, reduced smoking lowers morbidity and increases life expectancy. The net effect remains unclear, and is surrounded by a lot of uncertainty.

Our aim in this paper is to model and predict the future development of health in the United States, as well as the degree of uncertainty regarding the future development. We first apply the stochastic approach proposed by Lee and Carter (1992). This is a parsimonious modeling approach that consists of decomposing (in its original form) mortality into an age and a time effect. Such an approach seems to be relevant for modeling health as well, since health as a function of time and age shows similarities to mortality as a function of time and age. Importantly, the Lee and Carter (1992) model explicitly allows for quantifying the uncertainty surrounding the health development and its forecasts.

There are many studies forecasting health using a deterministic approach (e.g., Singer and Manton (1998), Jacobzone (2000), Jagger, Matthews, Spiers, Brayne, Comas-Herrera,

Robinson, Lindesay, and Croft (2006), Manton, Gu, and Lamb (2006a), and Manton, Lamb, and Gu (2007), to name just a few). Forecasting health using a deterministic approach has a main disadvantage that the forecast uncertainty is not quantified. The deterministic forecast might over- or underestimate the future development, even to large extent, and does not provide any information on such forecast errors. As a consequence, policy measures based on deterministic forecasts might turn out to be misguided, if the forecast error turns out to be substantial and the possibility of such a forecast error is not taken into account. See, for example, also Lee and Miller (2002) for a discussion of the drawbacks associated with deterministic forecasts. In response to these concerns, there is a growing body of literature on stochastic methods for demographic and health forecasting. Since the seminal work by Lee and Carter (1992), the stochastic approach to mortality forecasting has gained considerable attention, but is still less common for health modeling. Notable recent exceptions are Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012) and van Baal, Peters, Mackenbach, and Nusselder (2013). Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012) apply the Lee-Carter model to model and project transition probabilities between health states and death for the Dutch population at a disaggregated level. van Baal, Peters, Mackenbach, and Nusselder (2013) apply the Lee-Carter model to quantify the healthy life expectancy of different educational groups at an aggregated level. Our approach differs from theirs in the following way. We extend the Lee and Carter (1992) approach by including observed variables (GDP per capita and the unemployment rate), which turn out to capture most of the time variation in health. This has some advantages compared to the traditional Lee and Carter (1992) model. First, the model allows for a direct link to (economic) models that generate (future) scenarios of GDP and the unemployment rate. Second, it implies similar time variations in terms of health in countries with similar time variations in the development of GDP and unemployment. Finally, the model might generate more precise model-based forecasts, see also Niu and Melenberg (2014). The reason is that in the traditional Lee and Carter (1992) model, a latent variable is estimated that is assumed to capture the time variation. This estimated latent trend variable might result in some in-sample overfitting, when trying to capture the time variation as good as possible. Compared to GDP and unemployment, the latent estimated variable might be more volatile. This extra volatility is then translated into a wider forecast interval. In our extended Lee and Carter (1992) model, the included latent time variable will only capture the residual time variation, not yet captured by GDP and unemployment. If GDP and unemployment indeed capture most of the time variation, as we do find, then the possible overfitting by the included latent time variable likely will only be of minor importance. Another difference between our approach and that in Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012) is that we focus on modeling the time variation in aggregate health (using a time series of cross-sectional health status data), whereas they model the time variation in individ-

ual transitions in health status. While our approach has the disadvantage that it yields less detailed information regarding health and its relation to mortality, an important advantage is that we can use a much longer dataset. Whereas their time period covers 19 years (1989–2007), our study uses aggregated U.S. data over the period 1972-2010. The longer dataset might help to better capture long-term trends in health status at population level.

The remainder of the paper is organized as follows. In the next section, we formally define the health status index, and introduce the theoretical framework to estimate health changes stochastically. Next, in Section 5.3, we describe the health data and macroeconomic variables included in the study. Section 2.4 presents the estimation results on modeling the health dynamics for the United States from 1972 to 2010, distinguishing males and females. We then discuss the forecast of health changes in Section 2.5. Section 2.6 provides a sensitivity analysis. We conclude in Section 5.5.

2.2 Health modeling

In this section, we first present the health measurement used in this paper, focusing on the construction of the Health Status Index (HSI). Next, a latent framework is illustrated to model dynamic changes in the health process. We then extend the latent model by including observed macroeconomic information.

2.2.1 Health measurement

The analysis in this paper uses self-assessed health. Although there are some well-known drawbacks to using self-assessed health (such as, e.g., its subjective nature, possible biases, and heterogeneity), self-assessed health is a commonly used measure of health. While it is indeed subjective in nature, it can incorporate a variety of features of health, including not only physical aspects, but also cognitive and emotional health. Several studies show that it might provide useful information regarding an individual's working eligibility, health service demand, and long-term care needs, see, for instance, Branch, Jette, Evashwick, Polansky, Rowe, and Diehr (1981), Peng, Ling, and He (2010), and McGarry (2004).

In line with the health definition introduced by Imai and Soneji (2007a), we define the Health Status Index (HSI), $\pi_{x,t}$, to represent the proportion of the population of group x at time t with a certain health condition, for example, "good" or "bad." This Health Status Index, reflects the overall health level of the population of a certain age and at a certain time. In our application, the population consists of all males or all females in the U.S., and a group x consists of all individuals in this population at age x, with $x \in \{0, 1, \cdots, 85+\}$.

2.2.2 Health modeling in a latent framework

In this section, we model the development of the Health Status Index (HSI) over specific groups and time employing the original Lee and Carter (1992) framework, which is a parsimonious and latent modeling approach. The Lee-Carter model and its numerous extensions belong to the commonly used methods in mortality analysis. See, for instance, recent books by Girosi and King (2008) and Pitacco, Denuit, Haberman, and Olivieri (2009a), and references included in these works. Quantitative comparisons of the Lee-Carter model and its extensions can be found in, for example, Cairns, Blake, Dowd, Coughlan, Epstein, Ong, and Balevich (2007), Dowd, Cairns, Blake, Coughlan, Epstein, and Khalaf-Allah (2010), and Cairns, Blake, Dowd, Coughlan, Epstein, and Khalaf-Allah (2011). They conclude that no single model dominates all other models. Since our study is one of the first attempts to model health dynamics under a latent stochastic framework in the current literature, there is no reason to assume at this stage that a more complicated extension will outperform the original Lee-Carter framework.

Let $\pi_{x,t}$ denotes the health status index (HSI) of group x at time t. The Lee-Carter model assumes that some transformation F of $\pi_{x,t}$ satisfies the following relationship,

$$F(\pi_{x,t}) = \alpha_x + \beta_x \kappa_t + \epsilon_{x,t}, \tag{2.1}$$

where α_x describes the time-independent level of health as a function of x, κ_t is a time-dependent univariate latent variable, which represents the change in the overall level of $F(\pi_{x,t})$ over time, β_x describes the group-specific sensitivity to the overall level when κ_t varies, and $\epsilon_{x,t}$ is the error term, reflecting idiosyncratic time- and group-specific influences, with mean 0 and (possibly group-specific) variance $\sigma^2_{\epsilon,x}$.

In this model specification, α_x, β_x and κ_t are not uniquely identified. For instance, multiplying all β_x-s by a non-zero constant and dividing all κ_t-s by the same constant c, or adding a non-zero constant d to κ_t and subtracting $d \times \beta_x$ from α_x does not alter the systematic part of the model. Hence, Lee and Carter (1992) propose two normalization constraints,

$$\sum_t \kappa_t = 0 \text{ and } \sum_x \beta_x = 1. \tag{2.2}$$

The first constraint implies that for each x an estimate for α_x will be the average of the $F(\pi_{x,t})$ over time. The second one implies that β_x represents which fraction (over all groups) of the change in κ_t is captured by group x.[1] These normalizations identify the α_x-s and κ_t-s. The β_x-s are identified if the κ_t-process is not identically equal to zero. Thus, if we set $\beta_x = 0$ (all x) if $\kappa_t = 0$ (all t), then also the β_x-s are identified.

[1] As argued by Cairns, Blake, Dowd, Coughlan, Epstein, Ong, and Balevich (2007) and Pitacco, Denuit, Haberman, and Olivieri (2009a), the first constraint is a natural choice, whereas other choices of the second constraint have no impact on the quality of the fit, nor the model forecasts. Other constraints can be found in the literature, for instance, Wilmoth (1993) employs $\sum_x \beta_x^2 = 1$.

Originally, Lee and Carter (1992) use $F(z) = \log z$ when the dependent variable of interest is $m_{x,t}$, the central mortality death rate of group x at time t. As a benchmark, we also adopt the log-transformation of the HSI, though in case of $\pi_{x,t}$ other transformations might work better. In Section 2.6.1, we consider alternatives and show that the log-transformation seems to be a reasonable choice.

2.2.3 Lee-Carter model with observed variables

In the original Lee and Carter (1992) model, the latent κ_t captures the time trend. In this section, we introduce an extension of the Lee and Carter (1992) model, by including observed economic variables. Such an extension might help to better understand a possible trend in health, since the observed variables might capture some or even all of the trend instead of κ_t. Let Z_t be an m-dimensional vector containing as components of observed variables. Examples of Z_t can be macroeconomic variables (in our case logarithm of GDP and unemployment rate), or, alternatively, life-style related factors, such as alcohol and tobacco consumption (see Section 2.6.3 on the sensitivity analysis). The health curve is then modeled as

$$\log(\pi_{x,t}) = \alpha_x + \beta_x \kappa_t + \rho_x' Z_t + \epsilon_{x,t}, \tag{2.3}$$

where $\rho_x = \left(\rho_x^1, \cdots, \rho_x^m\right)'$ is an m-dimensional group specific parameter vector, containing the coefficients corresponding to Z_t. We normalize the components of the vector Z_t such that they have mean zero and variance one. However, adding some component of Z_t to κ_t and subtracting β_x from the corresponding component of ρ_x does not alter the systematic part of the model. Therefore, for identification purposes, we impose a constraint on ρ_x,

$$\sum_x \rho_x^i = 1, \text{ for each } i = 1, \ldots, m. \tag{2.4}$$

Suppose we observe $\pi_{x,t}$ and Z_t for $t \in \{t_1, \cdots, t_n\}$. If $\kappa = (\kappa_{t_1}, \cdots, \kappa_{t_n})'$ is not linearly dependent of the columns of $Z = (Z_{t_1}, \cdots, Z_{t_n})'$, then the β_x-s and ρ_x-s are identified. Thus, if we set $\beta_x = 0$ (all x), in case κ is linearly dependent of the columns of Z, then also the β_x-s and ρ_x-s are identified. See the appendix for a proof.

We estimate the model using the Newton-Raphson procedure, generalizing Renshaw and Haberman (2006), see the Appendix for details. Following Lee and Carter (1992), the estimated κ_t are adjusted by finding the value of κ_t for which the actual and expected total number of people who are in a certain health condition in each year are

equal, namely, we solve for $\widehat{\kappa}_t$ such that[2]

$$\sum_x H_{x,t} = \sum_x N_{x,t} \exp(\widehat{\alpha}_x + \widehat{\beta}_x \widehat{\kappa}_t + \widehat{\rho}_x' Z_t). \tag{2.5}$$

In addition, as we usually do not expect an irregular pattern of people's health changes with respect to group x, the age dependent estimates are smoothed using a spline method, proposed by Currie, Durban, and Eilers (2004), to fit the health surface.

Finally, to quantify the real trend in health captured by Z_t, we shall consider replacing the estimated $\widehat{\kappa}_t$ by $\tilde{\kappa}_t$ and the estimated ρ_x by $\tilde{\rho}_x$, where $\tilde{\kappa}_t$ is constructed such that it is orthogonal to Z_t, i.e,

$$\begin{aligned} \tilde{\kappa}_t &= \widehat{\kappa}_t - Z_t'(Z'Z)^{-1}(Z'\widehat{\kappa}), & (2.6)\\ \tilde{\rho}_x &= \widehat{\rho}_x + (Z'Z)^{-1}(Z'\widehat{\kappa})\beta_x, & (2.7)\\ \tilde{\beta}_x &= \widehat{\beta}_x. & (2.8) \end{aligned}$$

Since $\tilde{\kappa}_t$ by construction is orthogonal to Z_t, the resulting $\tilde{\rho}_x$ can be interpreted as capturing the "full" effect of Z_t on health. Moreover, $\tilde{\rho}_x = 0$ if Z_t would not have any effect and $\tilde{\kappa}_t = 0$, if there would be no remaining time effect next to Z_t.[3]

2.3 Data description

In this section, we describe the U.S. self-assessed health data and the macroeconomic variables used in this study.

2.3.1 Health data

The empirical analysis in this paper is based on consecutive annual cross-sectional self-assessed health data over the period 1972-2010 in the United States. The health data is obtained from the Integrated Health Interview Series (IHIS).[4] The IHIS documents the integrated self-assessed health of the civilian, non-institutionalized U.S. population, surveyed by the National Health Interview Survey (NHIS). The NHIS is a cross-sectional household face-to-face interview survey. It is conducted by the National

[2]The identification constraints will be satisfied by replacing $\widehat{\kappa}_t$ with $\widehat{\kappa}_t - \bar{\widehat{\kappa}}_t$ and $\widehat{\alpha}_x$ by $\widehat{\alpha}_x + \widehat{\beta}_x \bar{\widehat{\kappa}}_t$.

[3]Niu and Melenberg (2014) use this way of estimating their model for mortality, similar to (3.4), but then for mortality instead of health and with only GDP per capita as observed variable included. Their normalization is that κ is orthogonal to the space spanned by Z, instead of our normalization that κ is linearly independent of the columns of Z. Moreover, if $\kappa = 0$ then they set $\beta_x = 0$ (all x), while we set $\beta_x = 0$ (all x) if κ is linearly dependent of the columns of Z. These two ways of identifying the parameters are equivalent.

[4]Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 4.0. Minneapolis: University of Minnesota 2011. For general information, see `http://www.ihis.us/ihis/`.

Center for Health Statistics (NCHS) and Centers for Disease Control and Prevention (CDC). On average, around 42,000 households are interviewed annually since 1972. These households contain on average around 100,000 people. The annual response rate of the eligible households is close to 93%.[5] All household members are interviewed, with information of household members under age 18 provided by a knowledgeable adult member of the household. The annual average conditional persons' response rate on the self-assessed health variable is 99.5%.[6] Non-respond persons are people who refused, reported not ascertained or unknown. Detailed information on the household response rates and the conditional persons' response rates each year from 1972 to 2010 is provided in Table 2.3 in the appendix. In addition, the IHIS constructs a variable, person weight, representing the inverse probability of persons selected into the sample. The person weight is based on the Final Annual Weight in the original NHIS public use files and adjusted for non-response with post-stratification adjustments for age, race/ethnicity, and sex using the Census Bureau's population control totals.[7]

The NHIS survey rates an individual's health on a four-point scale (excellent, good, fair, or poor) for 1972-81 and a five-point scale (excellent, very good, good, fair, or poor) from 1982 until now. We define the health status index in the way that people are classified to be healthy unless they report "poor" or "fair." Accordingly, we define $H_{j,x,t} = 1$ when respondent j belonging to age group x in year t reports a "bad" health condition ("poor" or "fair"), and $H_{j,x,t} = 0$ otherwise. The Health Status Index of age class x in year t ($\pi_{x,t}$) is estimated as follows, using the IHIS constructed variable person weight ($\omega_{j,x,t}$) to make the Index representative for the U.S. population,

$$\widehat{\pi}_{x,t} = \frac{1}{\sum_{j=1}^{N_{x,t}} w_{j,x,t}} \sum_{j=1}^{N_{x,t}} w_{j,x,t} H_{j,x,t}, \tag{2.9}$$

where $N_{x,t}$ denotes the number of persons in age class x in year t.

In our analysis, we use data over the period 1972-2010 on males and females separately, where the groups are age classes ranging from age 0 to age 85+, where the age class 85+ consists of the individuals of age 85 and higher.[8] We exclude individuals with response "unknown", which is only a very small proportion of the entire sur-

[5]The NHIS reports that non-response households are those were not interviewed due to reasons including refusal, no one is home after repeated contact attempts, unacceptable partial interviews, or other reasons for no interview.

[6]The conditional persons' response rate is the ratio of the number of interviewed persons who provide health information to the number of interviewed persons.

[7]See `https://www.ihis.us/ihis/userNotes_weights.shtml`

[8]The variable "Health status," downloaded from the website `https://www.ihis.us/ihis-action/variables/group/health_general`, presents aggregated information from 1996 onwards (1996: 90+, 1997 and later: 85+. For this reason we also aggregate the data over the ages 85+ for the years 1972–1995.

vey sample.[9] The IHIS reports that the relative frequency of responses more favorable than "fair," i.e., combining "excellent," "very good," and "good" versus combining "excellent" and "good," is similar before and after 1982. This motivates our choice of constructing the health status index, with the aim to avoid a systematic shock because of the change of reported health categories.[10]

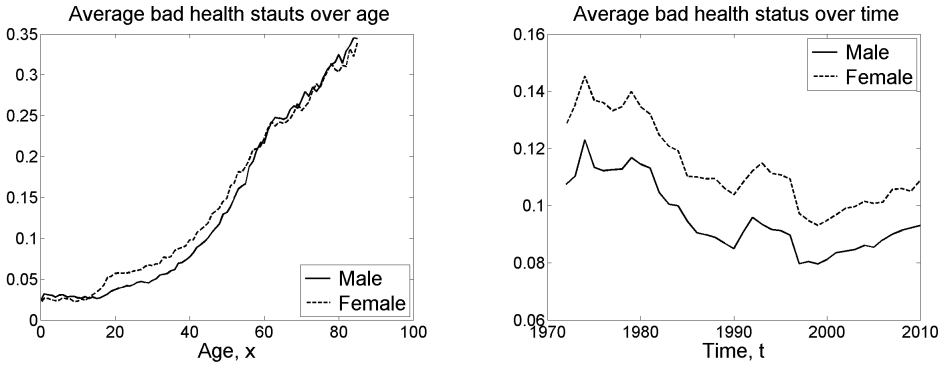

Figure 2.1 – Description of the Health Status Index in the U.S.

Data Source: Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 4.0. Minneapolis: University of Minnesota, 2011.
Note: The left graph shows the average bad health condition as a function of age averaged over time. The right graph shows the average bad health condition as a function of time averaged over age.

Figure 3.3 describes the average health status index over age (left graph) and over time (right graph) for both males and females.[11] As our health status index represents people's "bad" health, its growing patterns over age are expected. This indicates that, in general, people's health condition is getting worse as people age. Over time, we first see a decreasing and then a slightly increasing trend, implying that health changes are not just trended in one direction as time goes on.

2.3.2 Observed variables

It is well documented that population health is associated with the macroeconomic condition (see, e.g., Toffolutti and Suhrcke (2014), García-Muñoz, Neuman, and Neuman (2014), Baird, Friedman, and Schady (2011), Ruhm (2000), and Harvey Brenner (1979), to name a few). For instance, García-Muñoz, Neuman, and Neuman (2014)

[9] For males 0.53% and for females 0.54% are unknown.

[10] As part of the sensitivity analysis, we investigate whether there are systematic differences when using the whole sample, or only the subsample 1982–2010.

[11] The average health status index over age is calculated based on the total number of respondents among all the survey years. Similarly, the average health status index over time is calculated based on the total number of respondents at all ages.

find that self-reported health is largely affected by GDP per-capita. Ruhm (2004) and Ruhm (2003) suggest that higher income reduces the risk of morbidity and functional limitations. As for the effect of unemployment on health, the literature presents mixed evidence. On the one hand, a higher unemployment rate may result in reduced income and the loss of health insurance (see Cawley, Moriya, and Simon (2011)). This happens particularly in a country like the United States, in which workers receive health insurance coverage as employee benefits. The U.S. Census Bureau report Employment-Based Health Insurance: 2010[12] states that 56.5% of the U.S. population in 2010 relied on employment-based health insurance. This means that many working adults will lose health insurance once unemployed, and hence, will have limited access to healthcare (see Quinn, Catalano, and Felber (2009) and Catalano and Satariano (1998)). Consistent with that, Tefft and Kageleiry (2014) find that preventive healthcare decreases when unemployment increases. On the other hand, as argued by e.g., Ruhm (2000), unemployment might also positively affect people's health. This could occur, for example, when unemployment reduces job stress, or allows for more leisure and healthy behavior. Given this empirical evidence, we investigate whether GDP and unemployment rate can capture part of the trend in health.

We obtain these two macroeconomic variables from the Organisation for Economic Cooperation and Development (OECD) Statistics Extracts (the Country Statistical Profiles, 2010). The sample period is 1972-2010. GDP per capita is in real terms corrected by the inflation based on the year 2000. The in-sample evolutions of these two variables' are presented in Figure 3.5. Over the past 39 years, GDP per capita has a generally increasing trend, while the unemployment rate clearly fluctuates over time, with clear upward peaks around 1975, around 1982-1983, around 1992, around 2003, and around 2010. We shall examine whether these macroeconomic quantities will help to capture the trend in health, in addition to the latent time variable in the Lee-Carter model.[13]

[12] Report is available from the website `http://www.census.gov/prod/2013pubs/p70-134.pdf`

[13] In the sensitivity analysis in Section 2.6.3, we also investigate the performance of two life-style related variables, alcohol and tobacco consumption.

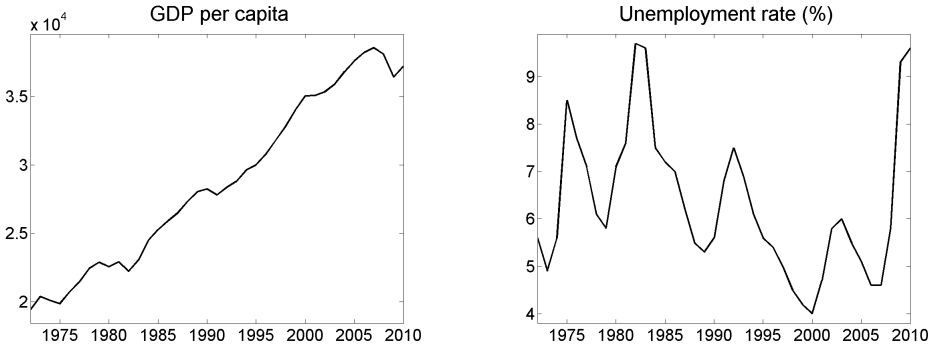

Figure 2.2 – Description of macroeconomic variables.

Note: The left graph describes the real GDP per capita in dollars, corrected by inflation. The right graph describes the total unemployment rate, as a fraction of the total labor force.

2.4 Model estimation

In this section, we present first the estimation results of the original Lee-Carter model (subsection 2.4.1), and then the results of its extended version with GDP and unemployment rate included (subsection 2.4.2).

2.4.1 Modeling health using the Lee-Carter model

In this subsection, we present the estimation results of the original Lee-Carter model for health, see equation (2.1). Figures 2.3 and 2.4 show the estimates for males and females, respectively. Each figure contains four panels. The upper left panel shows the estimated $\widehat{\alpha}_x$, the upper right panel the estimated $\widehat{\beta}_x$, the lower left panel the estimated $\widehat{\kappa}_t$ adjusted according to (2.5), and the lower right panel the estimated residuals. As irregular shapes of the estimated $\widehat{\alpha}_x$ and $\widehat{\beta}_x$ across age groups are usually not expected, we in addition show the smoothed estimates using B-splines, see Currie, Durban, and Eilers (2004).

For both males and females, besides the first 15 years in life, the increasing shape of estimated $\widehat{\alpha}_x$ (upper left panels) indicates that on average people's health is getting worse as people age. The estimated $\widehat{\kappa}_t$-s (left lower panels) are first declining, but then slightly trending up, indicating that the proportion of people in bad health has a decreasing trend over time, except for the last 10 years. Furthermore, the estimated $\widehat{\beta}_x$-s (upper right panels) show that the young are more sensitive to the time trend than the elderly. A simple and quick visual check of the model validity is to see whether the estimated residuals follow more or less a random pattern. In the lower right panels, the estimated residuals for both males and females indeed do not show a clear system-

atic structure, looking reasonably random. Nevertheless, for both males and females there seems to be a "line" separating the 1972–1981 period from the 1982–2010 period, suggesting a break between these subperiods. This likely corresponds to the survey design changes from the four-point to the five-point scale of individual health report since 1982. In Section 2.6.2, we present the estimation results for the subperiod 1982–2010, and show that there are no systematic differences between the whole sample period and this subperiod.

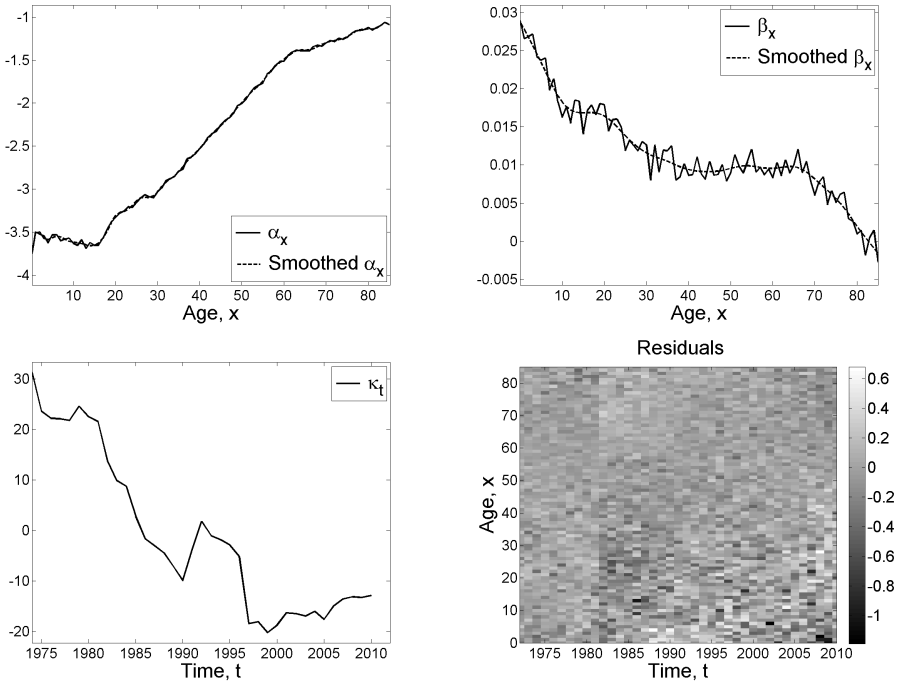

Figure 2.3 – Estimates of the Lee-Carter model for males.

Note: The upper left panel shows the non-smoothed and smoothed $\hat{\alpha}_x$. The upper right panel shows the non-smoothed and smoothed $\hat{\beta}_x$. The lower left panel shows $\hat{\kappa}_t$. The lower right panel shows the estimated residuals.

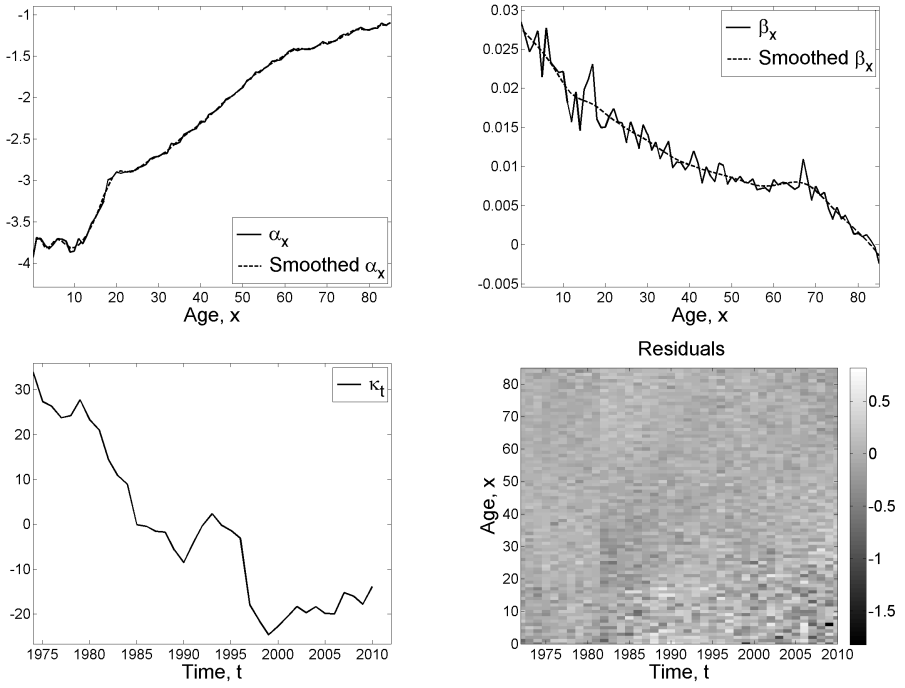

Figure 2.4 – Estimates of the Lee-Carter model for females.

Note: The upper left panel shows the non-smoothed and smoothed $\hat{\alpha}_x$. The upper right panel shows the non-smoothed and smoothed $\hat{\beta}_x$. The lower left panel shows $\hat{\kappa}_t$. The lower right panel shows the estimated residuals.

2.4.2 Modeling health with macroeconomic variables

In this section, we present the estimation results of the Lee-Carter model including the two macroeconomic variables, namely, GDP per capita in logarithmic form and the unemployment rate. The estimates of α_x and β_x are quite similar to the original Lee and Carter model, as presented in subsection 2.4.1, and therefore not reported. The plots of the residuals (also not reported) again do not reveal systematic patterns (other than the line separating the pre-1981 from the post-1982 period). Figure 2.5 presents the estimated $\tilde{\rho}_x$. For both males and females, the estimated $\tilde{\rho}_x$-s of log GDP (see left panels) show a negative correlation between people's bad health condition and GDP, where this negative correlation is strongest for the young. Thus, GDP and good health are positively correlated. The estimated $\tilde{\rho}_x$ corresponding to the unemployment rate (see right panels) show a positive correlation between the bad health condition and unemployment for most age classes, except for the very young and the very old. Thus, unemployment correlates negatively with good health.

The estimated $\tilde{\kappa}_t$-s, shown in Figure 2.6, seem to be stationary. The Augmented Dickey-Fuller test suggests that $\tilde{\kappa}_t$-s do not have unit roots for both genders. If $\tilde{\kappa}_t$-s are indeed stationary, this would imply that the trend in health is fully captured by the

macroeconomic fluctuations.

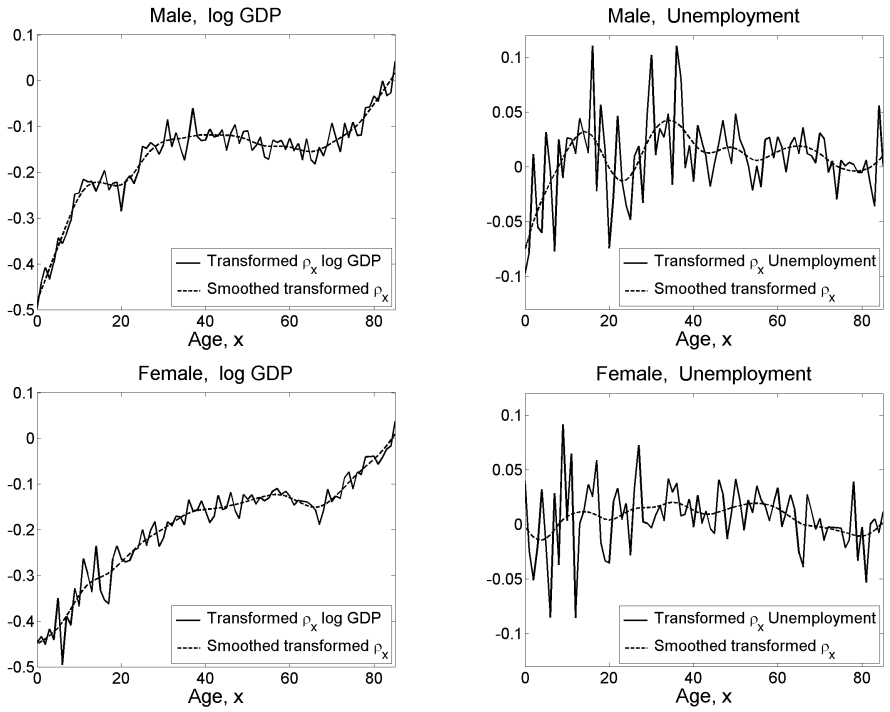

Figure 2.5 – Transformed ρ_x (i.e., $\tilde{\rho}_x$) in the extended Lee-Carter model.

Note: $\tilde{\rho}_x$-s of log GDP (left panels) and unemployment rate (right panels). The upper panels are for males. The lower panels are for females.

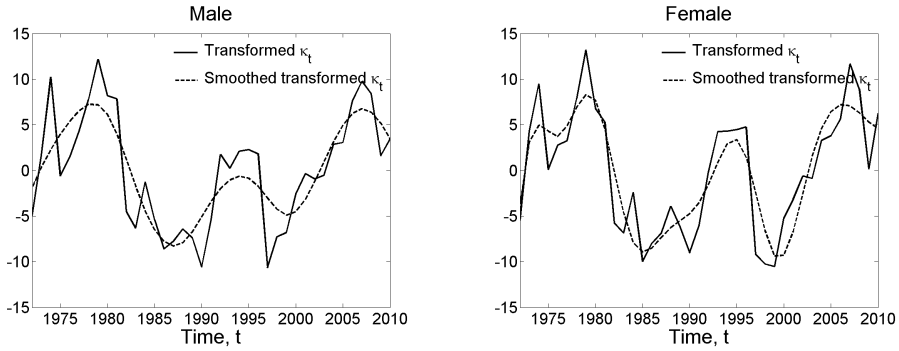

Figure 2.6 – Transformed κ_t (i.e., $\tilde{\kappa}_t$) in the extended Lee-Carter model.

Note: The left graph is for males. The right graph is for females.

Furthermore, to quantify the estimation inaccuracy, we use the bootstrap method, see the Appendix. Figure 2.7 shows the 95% confidence intervals for the smoothed $\tilde{\rho}_x$-s of log GDP (left panels) and unemployment rate (right panels) for both genders, calculated using 2000 bootstraps and allowing for age-specific $\hat{\sigma}^2_{\epsilon,x}$. For both males and females, at most of the ages, the GDP has a significant effect on health, since the

confidence intervals do not include zero, except at very high ages. However, the unemployment rate does not play a significant effect for many ages. On the other hand, the test results of the null hypothesis $H_0 : \tilde{\rho} = 0$ show that the included variables jointly have a significant effect on people's bad health.[14]

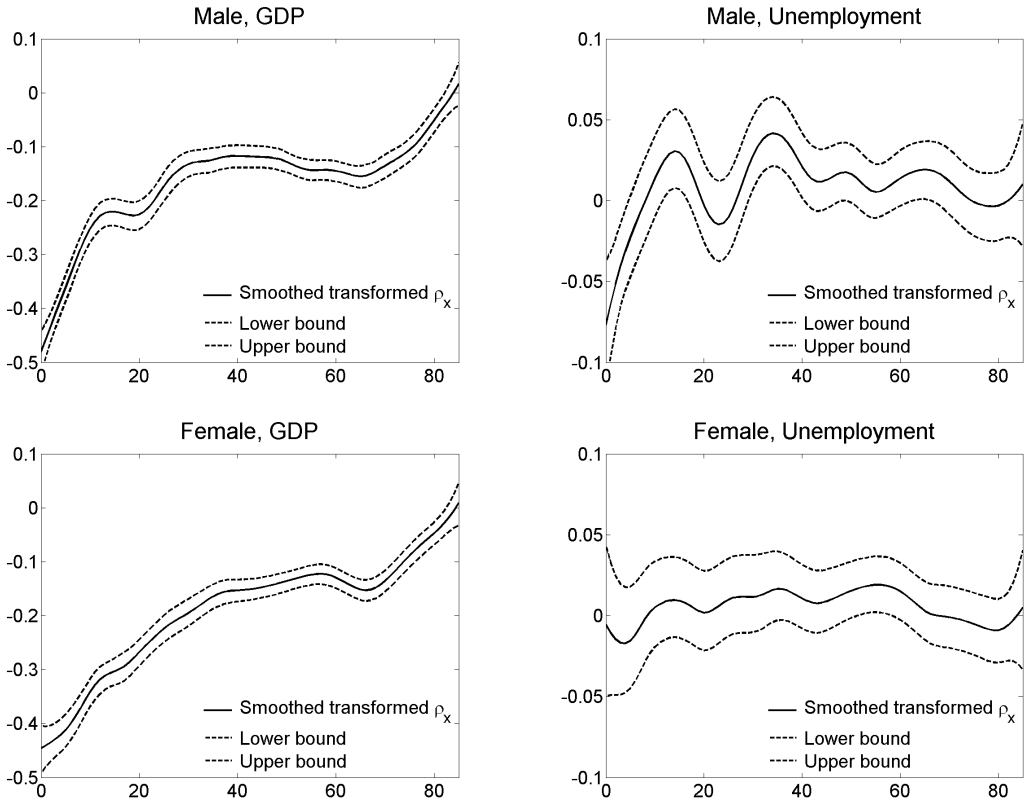

Figure 2.7 – Confidence intervals for smoothed $\tilde{\rho}_x$ in the extended Lee-Carter model.

Note: Left panels: log GDP. Right panels: unemployment rate. Upper panels: males. Lower panels: females.

We then compare the model fit for the two models of interest based on the Mean Square Errors (MSE). Results are presented in Table 2.1. We find that, compared with the original Lee-Carter model, the extended Lee-Carter model reduces the MSE-s by 18.0% for males, and by 19.9% for females. This leads to the conclusion that the Lee-Carter model with GDP and unemployment rate included yields a significant improvement in the model fit. Moreover, we also compare the values of the Bayes Information Criterion (BIC). In general, a smaller BIC value is preferred. This means that extra parameters are only included when there is a significant quality improvement of fit.

[14]Indeed, for males, we find for GDP as test statistic 85073 and for unemployment 274. For females, we find for GDP as test statistic 78779 and for unemployment 128. These are all significant at the conventional significance levels.

We see that the Lee-Carter model with GDP and unemployment rate also provides the smallest BIC values.

Table 2.1 – Comparison of model fit

	Male		Female	
	MSE(10^{-4})	BIC	MSE(10^{-4})	BIC
Original Lee-Carter model	5.158	-6.945	4.193	-7.153
Lee-Carter model with observed variables	4.228	-7.144	3.358	-7.375

2.5 Forecasting Health

Having developed and estimated the health model, we are now ready to consider forecasting health. The forecasting performance of a model is an important model evaluation criterion. In this section, we first address the method to forecast κ_t (for males and females) and observed variables (log GDP and unemployment rate). Based on the forecasts of these "independent" variables, the health status index is then forecasted using both the original Lee-Carter model and the Lee-Carter model extended with the macroeconomic variables.

In the traditional Lee-Carter approach applied to mortality data, the estimated κ_t is modeled and forecasted assuming an ARIMA(p,d,q) time series method. Lee and Carter (1992), and also many later applications, see Tuljapurkar, Li, and Boe (2000), conclude that the dynamics of κ_t in the mortality context can be described as a random walk with drift μ. This ARIMA(0,1,0) time series model is given by

$$\kappa_t = \mu + \kappa_{t-1} + e_t, \tag{2.10}$$

where the innovation e_t is assumed to follow a normal distribution with mean 0 and variance σ_e^2. However, in the Lee-Carter model with observed variables for health, we propose to apply models to describe the joint dynamic evolutions of κ_t (or $\tilde{\kappa}_t$) for males and females, and the observed variables. We consider three models, one using κ_t (males and females) and two using $\tilde{\kappa}_t$ (males and females). The latter two include one assuming $\tilde{\kappa}_t$ is stationary and one assuming $\tilde{\kappa}_t$ is nonstationary. We first describe the method of projecting κ_t. To indicate gender dependence we shall add a superscript $g \in \{m, f\}$. For example, κ_t^g is the κ_t for males if $g = m$ and the κ_t for females if $g = f$.

To undertake an out-of-sample analysis, we subdivide the dataset into an estimation period 1972-2000, and a forecasting period 2001-2010. Applying the Augmented Dickey-Fuller test, we find evidence that κ_t^m, κ_t^f, log GDP, and unemployment rate are all $I(1)$ processes, while their first differences are stationary over the estimation

period 1972-2000.[15] Therefore, we model the first differences $\Delta K_t \equiv K_t - K_{t-1}$ and $\Delta Z_t \equiv Z_t - Z_{t-1}$ in a Vector Auto Regression (VAR) model, which $K_t \equiv (\kappa_t^m, \kappa_t^f)'$. The Akaike Information Criterion (AIC) suggests a first order VAR model for $\boldsymbol{Y}_t$, as

$$\boldsymbol{Y}_t \equiv \begin{bmatrix} \Delta K_t \\ \Delta Z_t \end{bmatrix} = \boldsymbol{C} + \Theta \boldsymbol{Y}_{t-1} + \nu_t, \tag{2.11}$$

where $\boldsymbol{C}$ is a (4×1) parameter vector, Θ is a 4×4 coefficient matrix, and ν_t is a 4-dimensional vector of white noise terms with means zero and covariance matrix Σ_ν. Results of the VAR model estimation are shown in Table 2.4 in the Appendix. Using the VAR model, we are able to predict $\boldsymbol{Y}_{t+h}$, conditional on $\boldsymbol{Y}_t$ at time t. That is

$$\widehat{\boldsymbol{Y}}_{t+h} = (\boldsymbol{I} - \hat{\Theta})^{-1}(\boldsymbol{I} - \hat{\Theta}^h)\hat{\boldsymbol{C}} + \hat{\Theta}^h \boldsymbol{Y}_t,$$

where $\widehat{\boldsymbol{Y}}_{t+h} = (\widehat{\boldsymbol{K}}_{t+h}, \widehat{\boldsymbol{Z}}_{t+h})'$ denotes the h-period ahead forecast. Next, the forecast of $\log(\widehat{\pi}_{x,t+h})$ can be obtained from equation (3.4),

$$\log(\widehat{\pi}^g_{x,t+h}) = \widehat{\alpha}^g_x + \widehat{\beta}^g_x \widehat{\kappa}^g_{x,t+h} + (\widehat{\rho}^g_x)' \widehat{Z}_{t+h}, \tag{2.12}$$

where the superscripts $g \in \{m, f\}$ indicate the gender dependence.

Following the same principle, we employ two models to forecast health status index using $\tilde{K}_t = (\tilde{\kappa}_t^m, \tilde{\kappa}_t^f)'$. Assuming $\tilde{K}_t$ is nonstationary, we construct a VAR model, just like (4.12), but with $\Delta\tilde{K}_t$ instead of ΔK_t. As the final model, assuming $\tilde{K}_t$ is stationary, we use a first order Auto Regression (AR(1)) model for $\tilde{K}_t$ and a VAR model for Z_t, i.e. (4.12), but then restricted to ΔZ_t.

Due to the randomness of ν_t in equation (4.12), process risk arises. We quantify this process risk using simulation method as follows. Under the assumption $\nu_t \sim N(0, \Sigma_\nu)$, we simulate 2000 future innovations and sample paths from the multidimensional VAR model (4.12), then construct the corresponding forecasting intervals for the independent variables $\widehat{\boldsymbol{Y}}_{t+h}$ (using (4.12)) and the dependent variable $\widehat{\pi}_{x,t+h}$ (using (3.10)). Besides the process risk, due to the uncertainty caused by the inaccuracy of the estimated parameters $\hat{\boldsymbol{C}}$, $\hat{\Theta}$, and $\hat{\Sigma}_\nu$ in (4.12), as well as the uncertainty in the parameters $\hat{\alpha}_x$, $\hat{\beta}_x$, $\hat{\rho}_x$, $\hat{\kappa}_t$, and $\hat{\sigma}^2_{\epsilon,x}$ in (3.4), parameter risk arises. We further quantify the parameter risk by the bootstrap method, see the Appendix for further details.

[15] The Augmented Dickey-Fuller test statistics are (with p-values in brackets): log GDP in levels, $-0.03(0.63)$, and log GDP in first differences, $-2.88(0.01)$; unemployment in levels, $-1.60(0.10)$, and unemployment in first differences, $-4.17(0.001)$; κ_t^m in levels, $-0.64(0.41)$, and κ_t^m in first differences, $-4.97(0.001)$; κ_t^f in levels, $-0.44(0.48)$, and κ_t^f in first differences, $-4.41(0.001)$.

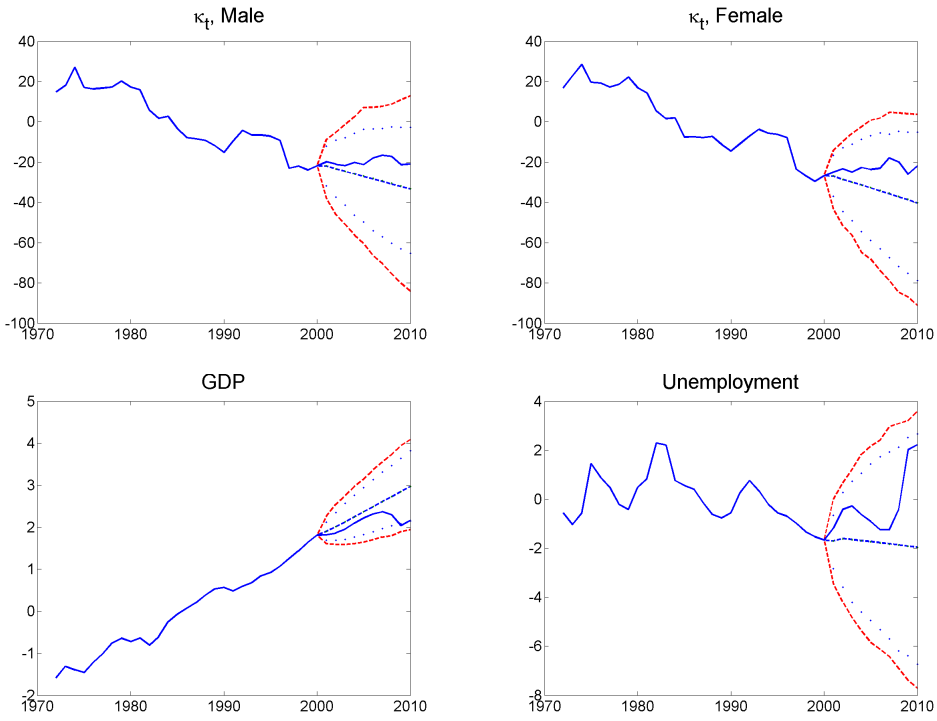

Figure 2.8 – Forecasts based on the VAR model.

Note: The upper panels show forecasts for κ_t^m (left panel) and κ_t^f (right panel). The lower panels show forecasts for GDP (left panel) and unemployment rate (right panel). Two confidence intervals are presented: narrower dotted curves present the uncertainty with process risk, wider dashed curves present the uncertainty with both process and parameter risks.

Figure 2.8 shows the forecasts of κ_t^m and κ_t^f (upper panels), and log GDP and unemployment rate (lower panels), together with their realized values (solid lines). The realized values of κ_t-s are constructed from the estimated κ_t-s for the whole sample 1972-2010. We rescale the estimated κ_t series for the sample 1972-2010 such that its value in the year 2000 is equal to the last estimated κ_t for the sample 1972-2000, and its summation from 1972 to 2000 is equal to 1. The rescaled κ_t series (1972-2010) is compared to the realized values. Indeed, in the upper panels, the dots, presenting the rescaled κ_t-s over the period 1972-2000, are very close to the estimated κ_t-s for the sample 1972-2000. 95% confidence intervals from process risk only, or both process and parameter risks in the *VAR* model during the out-of-sample period are shown.

Next, Figure 2.9 shows forecasts of the average health status index over time and age for males (upper panels) and females (lower panels), together with their realized values. In this figure, we also present forecasts with a jump-off bias correction, i.e., rescaling the forecast such that it starts off from the end-of-sample value without jump, see Lee and Miller (2001) for further details. In addition, the 95% confidence intervals from process risk only, or both process and parameter risks are presented in both fig-

ures. When quantifying the uncertainty in the parameters from (3.4), see left panels in Figure 2.9, we investigate two options, which either include or ignore the uncertainty in $\hat{\sigma}^2_{\epsilon,x}$. The results show that, on average, the out-of-sample observations in the forecasting period almost all fall into the constructed forecasted intervals.

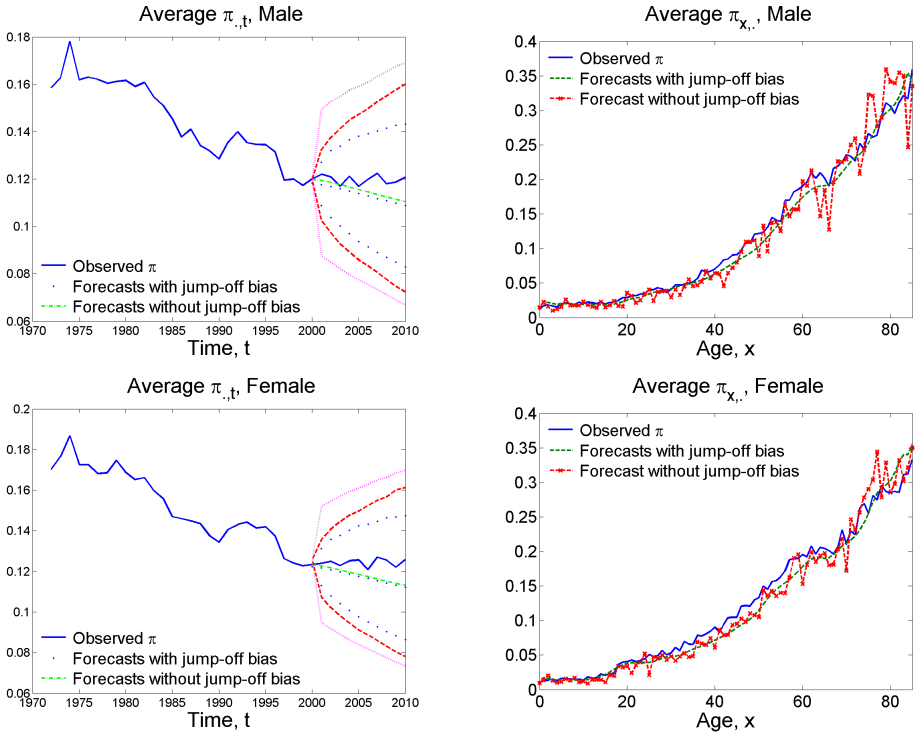

Figure 2.9 – Health forecasts based on the extended Lee-Carter model.

Note: Average forecasts of the bad health condition over age (left panels), and over time (right panels). The upper panels are for males. The lower panels are for females. Three confidence intervals are presented: narrowest dotted curves present the uncertainty with process risk, middle dashed curves present the uncertainty with both process and parameter risks, but excluding the uncertainty in $\hat{\sigma}^2_{\epsilon,x}$, and the largest intervals are for both process and parameter risks including the uncertainty in $\hat{\sigma}^2_{\epsilon,x}$.

To evaluate the forecasting performance of the models of interest, we use the mean squared forecasting error (MSFE), the mean absolute forecasting error (MAFE), and the mean forecast error (MFE), where we average the differences between the observations and the forecasts over both the age and the time dimensions. Table 2.2 presents the forecasting accuracy for males (the first panel) and females (the second panel). The first four rows of each panel show the forecast accuracy based on the Lee-Carter model, and the Lee-Carter models with the two macroeconomic variables employing three different variants, namely, the one with a VAR-model for ΔZ_t and an AR-process for $\tilde{K}_t$ ("$\tilde{K}_t$, AR"), the one with a VAR-model for both ΔZ_t and $\Delta\tilde{K}_t$ ("$\Delta\tilde{K}_t$"), and the one with a VAR-model for both ΔZ_t and ΔK_t ("ΔK_t"). The results show that, by including the observed variables, the MSFE and the MAFE are clearly improved compared with the

original Lee-Carter model, in particular for the VAR-model with K_t included ("ΔK_t"). In this case, the MSFE improves by 18.74% and 20.52% for males and females, respectively. Negative signs of the MFE indicate that on average we overforecast in all cases people's health improvement.

Table 2.2 – Comparison of forecast accuracy

	MSFE (10^{-4})	MAFE (10^{-2})	MFE (10^{-3})
	Male		
Original Lee-Carter	7.016	1.789	-3.191
$\tilde{K}_t$, AR	7.000	1.842	-9.860
$\Delta\tilde{K}_t$	5.721	1.613	-6.644
ΔK_t	5.701	1.613	-6.646
$\Delta K_t, Z_{t+h}$	6.347	1.650	-2.371
	Female		
Original Lee-Carter	7.179	1.914	-5.591
$\tilde{K}_t$, AR	6.595	1.871	-9.478
$\Delta\tilde{K}_t$	5.707	1.713	-7.503
ΔK_t	5.706	1.713	-7.503
$\Delta K_t, Z_{t+h}$	6.542	1.775	-3.949

Note: The first and the second panels are for males and females separately.
In the Lee-Carter model with GDP and unemployment rate (collected in Z_t),
"$\tilde{K}_t$ AR": predict with a VAR-model for ΔZ_t and an AR-process for $\tilde{K}_t$,
"$\Delta\tilde{K}_t$": predict with a VAR-model for both ΔZ_t and $\Delta\tilde{K}_t$,
"ΔK_t": predict with a VAR-model for both ΔZ_t and ΔK_t,
"$\Delta K_t, Z_{t+h}$": predict with the forecasted K_t (using the VAR model with ΔZ_t and ΔK_t) and the actually observed Z_{t+h}.

We examine an additional comparison with forecasts of the health status index using the realized values of the observed variables, i.e., we use equation (4.12) to forecast ΔK_t (untransformed), but in equation (3.10) we use the observed Z_{t+h}, instead of the forecasted $\widehat{Z}_{t+h}$ ("$\Delta K_t, Z_{t+h}$"). In this way, we eliminate the possible error due to estimating Z_t. However, we do not see an improvement of the forecasted π based on the realized values of the observed variables compared with the forecasts based on the forecasted values of the observed variables. What is noticeable is that our forecasting period includes the years of economic crisis. The large volatilities in changes of GDP and unemployment might be reduced if the forecasts are based on our VAR model. As a consequence, our VAR-forecasts for $\widehat{K}_{t+h}$ and $\widehat{Z}_{t+h}$ lead to better health forecasts, likely because health itself is a smooth process as well.

Finally, we construct a rolling window forecast to further test the forecasting power of the original Lee-Cater model and the Lee-Cater model with the macroeconomic variables, focusing on the untransformed κ_t. Based on the first fitting period, 1972-2000, we

compute 1 to 5 years ahead forecasts, 2001-2005, and determine the forecast errors by comparing the forecasts with the actual out-of-sample data. We then move the fitting period one year ahead, and compute again 1 to 5 years ahead forecasts, and the forecast errors. This procedure is repeated 6 times, until the last forecasting year is 2010. The lag order of the VAR model is chosen based on the AIC value in each rolling window estimation. According to the MSFE, we find quite a significant improvement of the forecasting performance from the Lee-Carter model with the two macroeconomic variables included compared with the original Lee-Carter model. Over the 6 rolling window forecasts, for males, the MSFE decreases on average 23.31%, with values between at most 28.19% and at least 20.13%, and for females, the MSFE decreases on average 21.45%, with values between at most 24.76% and at least 15.20%.

2.6 Sensitivity Analysis

This section presents a sensitivity analysis. We first examine whether a transformation of $\pi_{x,t}$ other than the log-transformation yields an improvement in the model fit. Next, we consider the subperiod 1982-2010, corresponding to the five-point scale of individual health report, instead of the whole period 1972–2010. Finally, we investigate two alternative life-style related factors, namely, alcohol and tobacco consumption, instead of, but also next to, the macroeconomic variables GDP per capita and the unemployment rate. We summarize most of the results.[16]

2.6.1 Different transformations of the health status index

Other transformations of $\pi_{x,t}$ than the log-transformation in equations (2.1) and (3.4) are possible. We experiment with alternative transformations $F(\pi_{x,t})$ to investigate whether these could increase the quality of the model fit. We consider the logit transformation ($F(\pi_{x,t}) = \log(\frac{\pi_{x,t}}{1-\pi_{x,t}})$), the Box-Cox transformation ($F(\pi_{x,t}) = \frac{\pi_{x,t}^a - 1}{a}$, given a certain parameter a, see Box and Cox (1964)), and the MacKinnon and Magee transformation ($F(\pi_{x,t}) = \frac{H(a\pi_{x,t})}{a}$, given a certain parameter a and with H the inverse hyperbolic sine transformation, see MacKinnon and Magee (1990)). We find that compared with the log-transformation, these alternative choices of $F(\pi_{x,t})$ do not result in a significant improvement of the mean square errors. Furthermore, they provide very similar estimates. Therefore, $F(\pi_{x,t}) = \log(\pi_{x,t})$ seems to be a good choice.

2.6.2 Analysis for subperiod 1982–2010

As reported in Section 2.4, Figures 2.3–2.4, for both males and females, there seems to be a "line" separating the 1972–1981 period from the 1982–2010 period, suggesting a

[16] Detailed results are available upon request from the corresponding author.

break between these two subperiods. This break subdivides the four-point scale data from the five-point scale data.

Therefore, we re-estimate the models, using the data of the subperiod 1982–2010 and further compare the estimates of the models using the whole sample period 1972-2010. Figure 2.10 shows some selected results. The left panels present both the estimated κ_t for the whole period 1972-2010 ($\widehat{\kappa}_t^{org}$) and the estimated κ_t for the subperiod 1982–2010 ($\widehat{\kappa}_t^{sub}$), where the latter is rescaled such that $\widehat{\kappa}_{t_1'}^{sub} = \widehat{\kappa}_{t_1'}^{org}$, and $\sum_{t=t_1'}^{T} \widehat{\kappa}_t^{sub} = \sum_{t=t_1'}^{T} \widehat{\kappa}_t^{org}$, with $t_1' = 1982$, the starting year of the subsample. For males (upper left panel) the estimated κ_t in both the whole and the subsample are quite close, whereas for females (lower left panel) the estimated κ_t in the subsample 1982–2010 seems to be somewhat more volatile than in the whole sample. The right panels show the estimated $\widehat{\beta}_x$ and its smoothed patterns for both males (upper right panel) and females (lower right panel). The results show that besides some increasing deviation between the subsample estimates and the whole sample estimates at the very young age groups, estimates in other age groups seem to be more or less similar. This deviation might be somewhat overemphasized by the smoothing method employed.

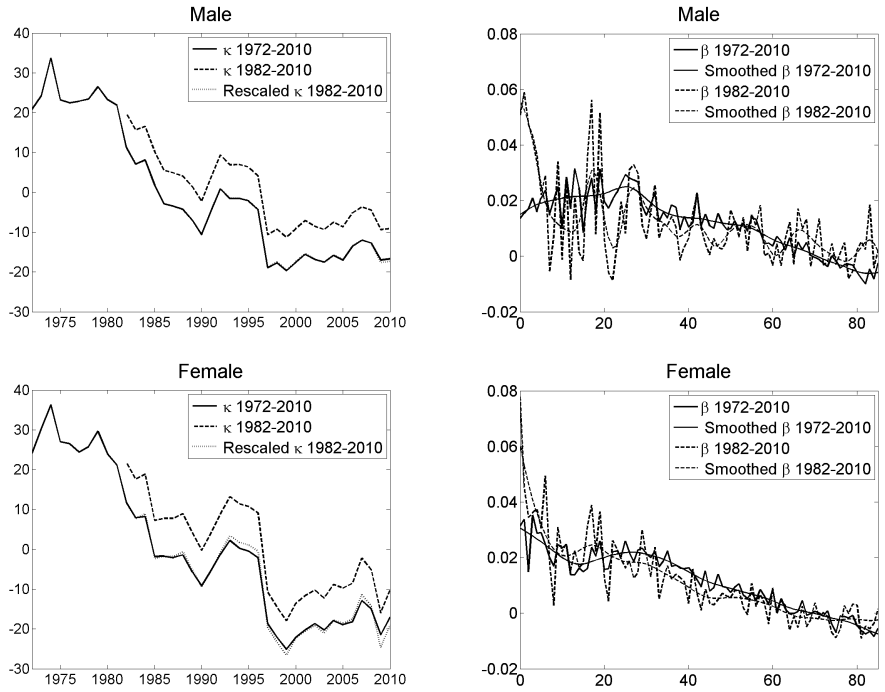

Figure 2.10 – Selected estimates of the Lee-Carter model for health, comparing 1982–2010 with 1972-2010.

Note: Upper panels: males. Lower panels: females. Left panels: κ_t. Right panels: β_x.

2.6.3 The choice of other observed variables

Population health is determined by many factors interactively. Besides the macroeconomic environment, there is extant evidence that health is affected by lifestyle choices. In our sensitivity analysis, we focus on alcohol consumption and smoking. According to Mokdad, Marks, Stroup, and Gerberding (2004) and McGinnis and Foege (1993), these lifestyle related factors are among the most important health risk factors in the United States. It is well-documented that smoking increases the risk of heart disease and lung cancer (the Center for Disease Control and Prevention (CDC).[17] Similarly, there is evidence of the health risks associated with alcohol consumption. It is argued that excessive alcohol use in the long term increases the risk of neurological, cardiovascular, and psychiatric problems, and can lead to lead to cancer and liver diseases[18] (see Corrao, Bagnardi, Zambon, and Vecchia (2004) and Rehm, Gmel, Sempos, and Trevisan (2003)). The World Health Organization (WHO (2011)) reports that almost 4% of the total deaths worldwide are caused by alcohol. This might be particularly relevant for the United States, because the average consumption of alcohol per person aged 15 years or older in the United States is higher than the average consumption worldwide (WHO (2011)). Therefore, we further investigate whether tobacco and alcohol consumption can capture trends in health. These two variables are obtained from the OECD Health Data (2010). Alcohol consumption is the annual consumption of pure alcohol in liters per person aged 15 years and over. Tobacco consumption is the annual consumption of tobacco items (for example, cigarettes, cigars) in grams per person aged 15 years and over. In the sample period 1972 to 2010, tobacco consumption has a steady decreasing trend, while alcohol consumption is increasing for the first 10 years, significantly decreasing in the following 10 years to a large extent, but then increasing again in the latest 15 years, although with a relatively small amount.

We estimate the Lee-Cater model with alcohol and tobacco consumption, instead of GDP and unemployment rate. We find that alcohol and tobacco consumption both have positive effects on people's "bad" health, reflected by the estimated transformed $\tilde{\rho}_x$. In addition, the estimated $\tilde{\rho}_x$ are jointly significantly different from 0. In terms of the mean square errors, the improvements compared with the original Lee-Carter model are 14.8% and 15.0% for males and females, respectively. However, these do not exceed the improvements when instead including GDP and unemployment rate (18.0% and 19.9%). This means that the two macroeconomic variables capture the health trend better than the two life-style related factors. The BIC values confirm this conclusion.[19]

Moreover, including three or all four observed variables reduces the mean square er-

[17] See `http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/`.

[18] `http://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm`

[19] The BIC values are now -7.089 for males and -7.300 for females, to be compared to -7.144 and -7.375, respectively.

rors considerably, but the corresponding BIC values are also much higher. Therefore, we consider the model with the two macroeconomic variables, GDP and unemployment rate included as the preferred one.

2.7 Conclusion

This paper develops a stochastic model to estimate and forecast health changes taking uncertainty into account. A better understanding of health dynamics is important for government policy decisions, such as the increase of the retirement age, or changes of the health expenditure. This article makes two main contributions. First, we consider the health dynamics as a stochastic process, and model it using the framework of Lee and Carter (1992). We find that the Lee-Carter model fits the self-assessed health data well for the United States. Second, we incorporate macroeconomic variables into the Lee-Carter model to better capture the health development in addition to the latent time factor. In this way, the health dynamics can be forecasted not only based on its historical pattern, but also on the basis of economy changes.

To summarize our key findings, first, a latent Lee-Carter framework works well to model health changes. Second, the Lee-Carter model with the macroeconomic variables leads to a significant improvement in the model fit. A large part of the time trend in health can be attributed to economic trends. Moreover, as suggested by the backtesting analysis, the Lee-Carter model with the macroeconomic variables significantly improves the accuracy for health forecasts compared with the original Lee-Carter model. We also conducted a sensitivity analysis. We first investigate various transformations of the health status index other than the log-transformation. We then experiment a subperiod analysis. Finally, alternative factors are also considered to capture the trend in health.

As alternative factors in our sensitivity analysis, we examined smoking and alcohol consumption. There are also other interesting factors, such as obesity. As reported by the CDC (Fryar, Carroll, and Ogden (2012)), the percentage of adults in the United States aged 20 years and over who are obese[20] increased from around 15% to 35% over the past 50 years. The National Institutes of Health (NIH (1998))[21] and Stanford Hospital & Clinics[22] report various potential obesity related health risks, including heart disease, diabetes, cancers, hypertension, stroke, liver and gallbladder diseases, etc. Thus, obesity is a serious risk factor, even becoming a more serious risk factor than tobacco (see Sturm (2002) and Mokdad, Marks, Stroup, and Gerberding (2004)).

[20] Here, somebody is classified as obese if the body mass index (BMI, kg/m^2) is larger than or equal to 30.

[21] See `http://www.nhlbi.nih.gov/guidelines/obesity/ob_gdlns.pdf`

[22] See `http://stanfordhospital.org/clinicsmedServices/COE/surgicalServices/generalSurgery/bariatricsurgery/obesity/effects.html`

We leave the investigation of the correlation between obesity and health for future research.

Appendix

Identification Lee-Carter model with observed variables

Let time $t = t_1, \ldots, t_n$, age $x = x_1, \ldots, x_k$, and $\dim(Z_t) = \dim(\rho_x) = m$, where $t_n \geq m + 2$. We write

$$\begin{aligned}\theta_a &= (\alpha_x, x = x_1, \ldots, x_k, \kappa_t, t = t_1, \ldots, t_n)\,,\\ \theta_b &= (\beta_x, x = x_1, \ldots, x_k, \rho_x, x = x_1 \ldots, x_k)\,,\\ \theta &= (\theta_a, \theta_b)\,,\end{aligned}$$

and

$$\begin{aligned}\ell\pi_{x,t} &= \log(\pi_{x,t}), x = x_1, \ldots, x_k, t = t_1, \ldots, t_n,\\ \ell\pi &= (\ell\pi_{x,t}, x = x_1, \ldots, x_k, t = t_1, \ldots, t_n)\,.\end{aligned}$$

Theorem

- If $\sum_t \kappa_t = 0$, $\sum_x \beta_x = 1$, and $\sum_x \rho_x^i = 1$, for each $i = 1, \ldots, m$, then

$$\theta_a^1 \neq \theta_a^2 \Rightarrow \ell\pi^1 \neq \ell\pi^2.$$

- Moreover, if

$$\begin{aligned}\boldsymbol{A} &\equiv \boldsymbol{A}(\tilde{t}_1, \tilde{t}'_1, \ldots, \tilde{t}_{m+1}, \tilde{t}'_{m+1})\\ &\equiv \begin{bmatrix} \kappa_{\tilde{t}_1} - \kappa_{\tilde{t}'_1} & Z_{\tilde{t}_1} - Z_{\tilde{t}'_1} \\ \vdots & \vdots \\ \kappa_{\tilde{t}_{m+1}} - \kappa_{\tilde{t}'_{m+1}} & Z_{\tilde{t}_{m+1}} - Z_{\tilde{t}'_{m+1}} \end{bmatrix} \in \mathbb{R}^{(m+1)\times(m+1)}.\end{aligned}$$

  is having a non-zero determinant, then

$$\theta^1 \neq \theta^2 \Rightarrow \ell\pi^1 \neq \ell\pi^2.$$

Thus, given the imposed normalizations, the α_x-s and κ_t-s are identified. Moreover, if $\kappa = (\kappa_{t_1}, \cdots, \kappa_{t_n})'$ is not linearly dependent of the columns of $Z = (Z_{t_1}, \cdots, Z_{t_n})'$, then the matrix $[\kappa \;\; Z]$ will have a $(m+2) \times (m+1)$ sub-matrix of rank $m+1$. Pre-multiplying this sub-matrix by a $(m+1) \times (m+2)$ differencing-operator of rank $m+1$ yields an $\boldsymbol{A}$ with non-zero determinant, implying that then also the β_x-s and ρ_x-s are identified.

Proof

- Assume $\alpha_x^1 \neq \alpha_x^2$. Then we have

$$\begin{aligned}\sum_t \log(\pi_{x,t}^1) &= \alpha_x^1 + \beta_x^1 \sum_t \kappa_t^1 + (\rho_x^1)' \sum_t Z_t \\ &\neq \alpha_x^2 + \beta_x^2 \sum_t \kappa_t^2 + (\rho_x^2)' \sum_t Z_t = \sum_t \log(\pi_{x,t}^2).\end{aligned}$$

- Assume $\alpha_x^1 = \alpha_x^2$ for all x and $\kappa_t^1 \neq \kappa_t^2$. Then we have

$$\begin{aligned}\sum_x \log(\pi_{x,t}^1) &= \sum_x \alpha_x^1 + \sum_x \beta_x^1 \kappa_t^1 + \sum_x (\rho_x^1)' Z_t \\ &\neq \sum_x \alpha_x^2 + \sum_x \beta_x^2 \kappa_t^2 + \sum_x (\rho_x^2)' Z_t = \sum_x \log(\pi_{x,t}^2).\end{aligned}$$

- Assume $\alpha_x^1 = \alpha_x^2$ for all x, $\kappa_t^1 = \kappa_t^2$ for all t, and $\begin{pmatrix} \beta_x^1 \\ \rho_x^1 \end{pmatrix} \neq \begin{pmatrix} \beta_x^2 \\ \rho_x^2 \end{pmatrix}$, and assume $\det(\boldsymbol{A}) \neq 0$ (for $\kappa_t = \kappa_t^1 = \kappa_t^2$), then we have

$$\begin{pmatrix} \ell\pi_{x,\tilde{t}_1}^1 - \ell\pi_{x,\tilde{t}_1'}^1 \\ \vdots \\ \ell\pi_{x,\tilde{t}_{m+1}}^1 - \ell\pi_{x,\tilde{t}_{m+1}'}^1 \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} \beta_x^1 \\ \rho_x^1 \end{pmatrix} \neq \boldsymbol{A} \begin{pmatrix} \beta_x^2 \\ \rho_x^2 \end{pmatrix} = \begin{pmatrix} \ell\pi_{x,\tilde{t}_1}^2 - \ell\pi_{x,\tilde{t}_1'}^2 \\ \vdots \\ \ell\pi_{x,\tilde{t}_{m+1}}^2 - \ell\pi_{x,\tilde{t}_{m+1}'}^2 \end{pmatrix}.$$

Estimation

We estimate the parameters α, β, ρ, and κ in (3.4) by solving iteratively the first order conditions resulting from minimizing

$$\mathcal{F}_{LS}(\alpha, \beta, \rho, \kappa) = \sum_{x=x_1}^{x_k} \sum_{t=t_1}^{t_n} (\pi_{x,t} - \alpha_x - \beta_x \kappa_t - \rho_x' Z_t)^2.$$

The system of equations to be solved iteratively is obtained by setting the partial derivatives of $\mathcal{F}_{LS}(\alpha, \beta, \rho, \kappa)$ with respect to α, κ, β, and ρ equal to zero:

$$\begin{aligned}0 &= \sum_{t=t_1}^{t_n} (\pi_{x,t} - \alpha_x - \beta_x \kappa_t - \rho_x' Z_t), x = x_1, \cdots, x_k \\ 0 &= \sum_{x=x_1}^{x_k} \beta_x (\pi_{x,t} - \alpha_x - \beta_x \kappa_t - \rho_x' Z_t), t = t_1, \cdots, t_n,\end{aligned}$$

$$
\begin{aligned}
0 &= \sum_{t=t_1}^{t_n} \kappa_t(\pi_{x,t} - \alpha_x - \beta_x\kappa_t - \rho_x' Z_t), x = x_1, \cdots, x_k, \\
0 &= \sum_{t=t_1}^{t_n} Z_t(\pi_{x,t} - \alpha_x - \beta_x\kappa_t - \rho_x' Z_t), x = x_1, \cdots, x_k.
\end{aligned}
$$

We solve this system by using iteratively the following univariate Newton-Raphson schemes (given some starting values):

$$
\begin{aligned}
\widehat{\alpha}_x^{(r+1)} &= \widehat{\alpha}_x^{(r)} + \frac{\sum_{t=t_1}^{t_n}(\pi_{x,t} - \widehat{\alpha}_x^{(r)} - \widehat{\beta}_x^{(r)}\widehat{\kappa}_t^{(r)} - \widehat{\rho}_x^{(r)'} Z_t)}{t_n - t_1 + 1}, \\
\widehat{\kappa}_t^{(r+1)} &= \widehat{\kappa}_t^{(r)} + \frac{\sum_{x=x_1}^{x_k} \widehat{\beta}_x^{(r)}(\pi_{x,t} - \widehat{\alpha}_x^{(r+1)} - \widehat{\beta}_x^{(r)}\widehat{\kappa}_t^{(r)} - \widehat{\rho}_x^{(r)'} Z_t)}{\sum_{x=x_1}^{x_k}(\widehat{\beta}_x^{(r)})^2}, \\
\widehat{\beta}_x^{(r+1)} &= \widehat{\beta}_x^{(r)} + \frac{\sum_{t=t_1}^{t_n} \widehat{\kappa}_t^{(r+1)}(\pi_{x,t} - \widehat{\alpha}_x^{(r+1)} - \widehat{\beta}_x^{(r)}\widehat{\kappa}_t^{(r+1)} - \widehat{\rho}_x^{(r)'} Z_t)}{\sum_{t=t_1}^{t_n}(\widehat{\kappa}_t^{(r+1)})^2}, \\
\widehat{\rho}_x^{(r+1)} &= \widehat{\rho}_x^{(r)} + \frac{\sum_{t=t_1}^{t_n} Z_t(\pi_{x,t} - \widehat{\alpha}_x^{(r+1)} - \widehat{\beta}_x^{(r+1)}\widehat{\kappa}_t^{(r+1)} - \widehat{\rho}_x^{(r)'} Z_t)}{\sum_{t=t_1}^{t_n} Z_t^2}.
\end{aligned}
$$

After each update, these parameters are adjusted by the normalization constraints (3.5) and (2.4). After convergence, $\widehat{\kappa}_t^{(r+1)}$ are further adjusted by fitting the total observed and expected number of people who are in bad health condition for each year t.

Table 2.3 – Description of the National Health Interview Survey

Year	Households			Persons								
	Eligible	Non-Response	Response Rate (%)	Interviewed			Non-Response			Response Rate (%)		
				Total	Male	Female	Total	Male	Female	Total	Male	Female
1972	46,149	1,785	96.13	132,891	63,995	68,896	1,074	526	548	99.19	99.18	99.20
1973	42,135	1,460	96.53	120,493	57,937	62,556	726	370	356	99.40	99.36	99.43
1974	41,314	1,294	96.87	116,287	55,707	60,580	647	304	343	99.44	99.45	99.43
1975	41,649	1,274	96.94	116,289	55,821	60,468	695	338	357	99.40	99.39	99.41
1976	41,559	1,554	96.26	113,178	54,118	59,060	505	235	270	99.55	99.57	99.54
1977	41,277	1,372	96.68	111,279	53,242	58,037	538	241	297	99.52	99.55	99.49
1978	41,164	1,580	96.16	109,940	52,578	57,362	577	256	321	99.48	99.51	99.44
1979	41,883	1,451	96.54	110,530	52,980	57,550	786	383	403	99.29	99.28	99.30
1980	39,226	1,148	97.07	102,629	49,101	53,528	340	149	191	99.67	99.70	99.64
1981	41,265	1,248	96.98	107,480	51,575	55,905	467	220	247	99.57	99.57	99.56
1982	39,988	1,194	97.01	103,923	49,812	54,111	791	362	429	99.24	99.27	99.21
1983	40,912	1,332	96.74	105,620	50,568	55,052	600	288	312	99.43	99.43	99.43
1984	41,471	1,475	96.44	105,290	50,335	54,955	457	222	235	99.57	99.56	99.57
1985	36,399	1,555	95.73	91,531	43,587	47,944	437	233	204	99.52	99.47	99.57
1986	24,698	860	96.52	62,052	29,532	32,520	264	117	147	99.57	99.60	99.55
1987	49,569	2,329	95.30	122,859	58,411	64,448	738	366	372	99.40	99.37	99.42
1988	50,061	2,576	94.85	122,310	58,037	64,273	736	356	380	99.40	99.39	99.41
1989	48,054	2,343	95.12	116,929	55,570	61,359	715	308	407	99.39	99.45	99.34
1990	48,680	2,204	95.47	119,631	57,134	62,497	628	304	324	99.48	99.47	99.48

Table 3 Continued: Description of the National Health Interview Survey

Year	Households			Persons								
	Eligible	Non-Response	Response Rate (%)	Interviewed			Non-Response			Response Rate (%)		
				Total	Male	Female	Total	Male	Female	Total	Male	Female
1991	48,853	2,092	95.72	120,032	57,324	62,708	710	359	351	99.41	99.37	99.44
1992	51,643	2,242	95.66	128,412	61,252	67,160	649	327	322	99.49	99.47	99.52
1993	44,978	1,971	95.62	109,671	52,467	57,204	477	212	265	99.57	99.60	99.54
1994	48,584	2,879	94.07	116,179	55,364	60,815	952	442	510	99.18	99.20	99.16
1995	41,824	2,585	93.82	102,467	48,809	53,658	1,190	543	647	98.84	98.89	98.79
1996	25,990	1,619	93.77	63,402	30,358	33,044	733	348	385	98.84	98.85	98.83
1997	43,370	3,538	91.84	103,477	49,561	53,916	538	244	294	99.48	99.51	99.45
1998	42,440	4,231	90.03	98,785	47,735	51,050	540	265	275	99.45	99.44	99.46
1999	42,882	5,309	87.62	97,059	46,943	50,116	442	207	235	99.54	99.56	99.53
2000	43,437	4,805	88.94	100,618	48,547	52,071	422	203	219	99.58	99.58	99.58
2001	43,797	4,865	88.89	100,760	48,689	52,071	417	205	212	99.59	99.58	99.59
2002	40,377	4,216	89.56	93,386	44,895	48,491	476	221	255	99.49	99.51	99.47
2003	40,266	4,345	89.21	92,148	44,504	47,644	470	219	251	99.49	99.51	99.47
2004	42,089	5,510	86.91	94,460	45,813	48,647	381	175	206	99.60	99.62	99.58
2005	44,540	6,031	86.46	98,649	47,732	50,917	290	139	151	99.71	99.71	99.70
2006	33,468	4,264	87.26	75,716	36,561	39,155	210	94	116	99.72	99.74	99.70
2007	33,615	4,349	87.06	75,764	36,726	39,038	160	74	86	99.79	99.80	99.78
2008	33,911	5,121	84.90	74,236	35,983	38,253	137	77	60	99.82	99.79	99.84
2009	41,177	7,321	82.22	88,446	42,875	45,571	102	53	49	99.88	99.88	99.89
2010	43,208	8,879	79.45	89,976	43,545	46,431	116	58	58	99.87	99.87	99.88

Quantifying parameter risk by the bootstrap method

To quantify the accuracy of the estimates, we apply the bootstrap method. For details about bootstrapping in the Lee-Carter framework, we refer to Pitacco, Denuit, Haberman, and Olivieri (2009a), in particular pages 229-232. First, we create a matrix $\boldsymbol{R}$ of residuals with components $\hat{\epsilon}_{x,t}$, given by

$$\hat{\epsilon}_{x,t} = \log(\pi_{x,t}) - \log(\hat{\pi}_{x,t}),$$

where $\log(\hat{\pi}_{x,t}) = \hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t + \hat{\rho}'_x Z_t$. From these, we generate $B = 2000$ replications $\boldsymbol{R}^b$ with components $\hat{\epsilon}^b_{x,t}$, $b = 1, 2, ..., B$, by sampling with replacement the components in the matrix $\boldsymbol{R}$. We assume the residuals are independently distributed and we allow for heteroskedasticity by means of age-specific variances $\sigma^2_{\epsilon,x}$. We then create the corresponding bootstrapped logarithm of the health status index $\log(\hat{\pi}^b_{x,t})$,

$$\log(\hat{\pi}^b_{x,t}) = \hat{\alpha}_x + \hat{\beta}_x\hat{\kappa}_t + \hat{\rho}'_x Z_t + \hat{\epsilon}^b_{x,t}.$$

Using the bootstrapped dependent variables, we re-estimate for each b the parameters, yielding the estimates $\hat{\alpha}^b_x$, $\hat{\beta}^b_x$, $\hat{\kappa}^b_t$, and $\hat{\rho}^b_x$. The confidence intervals for the estimates are derived from the corresponding bootstrapped percentiles.

To quantify the forecasting uncertainty in the parameters from both the Lee-Carter model with observed variables (3.4) and the VAR model (4.12), we again apply the bootstrap method and construct the bootstrap percentile intervals for the forecasts. The procedure is carried out as follows:

1. First, for each bootstrap $b = 1, 2, ..., B$, with $B = 2000$, we follow the same procedures as above and obtain the bootstrapped estimates $\hat{\alpha}^{b,g}_x$, $\hat{\beta}^{b,g}_x$, $\hat{\rho}^{b,g}_x$, $\hat{\kappa}^{b,g}_t$, and residuals $\hat{\epsilon}^{b,g}_{x,t}$ from the Lee-Carter model with observed variables (3.4). $g \in \{m, f\}$, where $g = m$ for males, and $g = f$ for females.

2. We then estimate the VAR model, see equation (4.12), for $\hat{\hat{\boldsymbol{Y}}}^b_t = (\Delta K^b_t, \Delta Z_t)'$, where $K^b_t = (\hat{\kappa}^{b,m}_t, \hat{\kappa}^{b,f}_t)'$. In turn, we estimate $\hat{\hat{\boldsymbol{C}}}^b$ and $\hat{\hat{\Theta}}^b$, and the residuals $\hat{\hat{v}}^b_t = \boldsymbol{Y}_t - \hat{\hat{\boldsymbol{Y}}}^b_t$, where $\hat{\hat{\boldsymbol{Y}}}^b_t = (\hat{\hat{\boldsymbol{C}}}^b + \hat{\hat{\Theta}}^b \boldsymbol{Y}_{t-1})$. This also results in the variance covariance matrix $\hat{\hat{\Sigma}}^b_v = \widehat{Cov}(\hat{\hat{v}}^b_t)$.

3. To ensure that the bootstrapped residuals $\hat{v}^b_t$ have the same variance covariance matrix $\hat{\hat{\Sigma}}^b_v$, the next procedure is followed. We first decompose $\hat{\hat{\Sigma}}^b_v = \boldsymbol{A}^b \boldsymbol{D}^b \boldsymbol{A}^{b'}$, where $\boldsymbol{D}^b$ is a unique diagonal matrix with positive entries along the principal diagonal and $\boldsymbol{A}^b$ is a unique lower triangular matrix with 1-s along the principal diagonal. We then construct $\hat{\hat{u}}^b_t = {\boldsymbol{A}^b}^{-1} \hat{\hat{v}}^b_t$. Next, $\hat{u}^b_t$ are generated by sampling with replacement from $\hat{\hat{u}}^b_t$. The bootstrapped residuals are in turn computed as $\hat{v}^b_t = \boldsymbol{A}^b \hat{u}^b_t$..

4. The bootstrapped series $\boldsymbol{Y}_t^b$ are obtained by $\boldsymbol{Y}_t^b = \hat{\hat{\boldsymbol{Y}}}_t^b + \hat{v}_t^b$. We estimate parameters $\hat{\boldsymbol{C}}^b$ and $\hat{\Theta}^b$ for $\boldsymbol{Y}_t^b$ in the *VAR* model.

5. Then we generate projections $\hat{\boldsymbol{Y}}_{t+h}^b$, based on bootstrapped estimates $\hat{\boldsymbol{C}}^b$ and $\hat{\Theta}^b$. The future errors $\hat{v}_{t+h}^b$ are sampled from a multivariate normal distribution with mean 0 and variance covariance matrix $\hat{\Sigma}_v^b$.

6. With the estimates in the Lee-Carter model $\hat{\alpha}_x^{b,g}$, $\hat{\beta}_x^{b,g}$, $\hat{\rho}_x^{b,g}$, projected $\hat{\kappa}_{t+h}^{b,g}$, and $\hat{Z}_{t+h}^b$, we can generate the projection $\hat{\pi}_{x,t+h}^{b,g}$ based on these bootstrapped estimates for both males and females. The future errors $\hat{\epsilon}_{x,t+h}^{b,g}$ are sampled from $\hat{\epsilon}_{x,t}^{b,g}$.

7. Finally, the confidence intervals of forecasted $\hat{\boldsymbol{Y}}_{t+h}$ are constructed based on the bootstrapped values $\hat{\boldsymbol{Y}}_{t+h}^b$. Similarly, the confidence intervals of forecasted $\hat{\pi}_{x,t+h}^g$ are obtained based on the bootstrapped values $\hat{\pi}_{x,t+h}^{b,g}$.

VAR-estimation results

Table 2.4 – Estimates of VAR model for males and females, Equation (4.12)

$$\boldsymbol{Y}_t \equiv \begin{bmatrix} \Delta K_t \\ \Delta Z_t \end{bmatrix} = \boldsymbol{C} + \Theta \boldsymbol{Y}_{t-1} + \nu_t$$

	$\boldsymbol{C}$	Θ			
		$\Delta\kappa^m_{t-1}$	$\Delta\kappa^f_{t-1}$	$\Delta\log(GDP)_{t-1}$	$\Delta UnEmp_{t-1}$
$\Delta\kappa^m_t$	$-2.21(1.90)$	$-0.48(0.41)$	$0.50(0.40)$	$9.50(15.49)$	$1.04(2.89)$
$\Delta\kappa^f_t$	$-2.50(2.00)$	$-0.15(0.43)$	$0.28(0.42)$	$11.19(16.26)$	$2.49(3.04)$
$\Delta\log(GDP)_t$	$0.10(0.04)$	$-0.01(0.01)$	$-0.002(0.01)$	$0.05(0.34)$	$-0.02(0.06)$
$\Delta UnEmp_t$	$0.31(0.21)$	$0.10(0.05)$	$-0.05(0.04)$	$-2.49(1.71)$	$-0.06(0.32)$

$$\boldsymbol{Y}_t \equiv \Delta K_t = \boldsymbol{C} + \nu_t$$

	$\boldsymbol{C}$
$\Delta\kappa^m_t$	$-1.38(0.83)$
$\Delta\kappa^f_t$	$-1.61(0.83)$

Note: The first panel presents the VAR model estimates when employing the Lee-Carter model with the macroeconomic variables. The second panel presents the VAR model estimates when employing the original Lee-Carter model.
Standard errors are provided in brackets.
$\Delta K_t = (\Delta\kappa^m_t, \Delta\kappa^f_t)'$, $\Delta Z_t = (\Delta\log(GDP)_t, \Delta UnEmp_t)'$, where $UnEmp$ denotes unemployment rate.

CHAPTER 3

DO AMERICANS LIVE LONGER AND HEALTHIER? FORECASTING HEALTHY LIFE EXPECTANCY BY INCLUDING DYNAMIC EVOLUTIONS OF MORTALITY, HEALTH, AND MACROECONOMIC VARIABLES

This Chapter is based on Yang, De Waegenaere, and Melenberg (2013a)

We study the evolution in remaining life expectancy and healthy life expectancy of the U.S. population from 1972 to 2020, distinguished by genders. We propose a methodology to forecast life expectancy and healthy life expectancy by incorporating the joint future developments of age-specific mortality and health prevalence rates, GDP, and the unemployment rate. We also quantify the uncertainties of (healthy) life expectancy taking into account the joint evolution of mortality, health, and macroeconomic variables. Finally, we make a comparison with the corresponding models without GDP and the unemployment rate or without taking into account the joint development of mortality and health.

3.1 Introduction

One of the remarkable changes in the United States over the past century is the population aging due to the baby boom and the unprecedented declines in mortality rates (Robine, Romieu, and Michel (2003)). The *U.S. National Vital Statistics Report, 2010* publish U.S. life expectancy for the total population by different age groups, using a period

life table. It reports that U.S. life expectancy at birth increases from 49.24 years in 1900 to 68.07 years by 1950, and to 77.7 in the second half of the century. The continuation of this pattern leads to a rapidly aging population. The accompanied increasing pension payment and health care spending attracts considerable attention from scholars, policymakers, and insurers. In addition to these growing concerns in the demography of aging, a question that is of great interest is how likely Americans are living longer with higher quality lives. Therefore, besides life expectancy, which measures the average remaining years of life a person can expect to live, healthy life expectancy (HLE) is increasingly recognized by researchers as a measure of the quality of life. It measures the expected remaining years of life in good health for persons at a given age and time. Estimates of healthy life expectancy (HLE) are available in many developed and developing countries (Jagger, Barberger-Gateau, and Robine (2005)) and focus on multiple dimensions of health, such as disability, physical functioning, disease prevalence, and self-assessed health (Crimmins, Hayward, Saito, and Kenkyūjo (1996) and Laditka and Laditka (2002)). Combined with life expectancy, healthy life expectancy has been used to study issues related to living longer and improving the quality of life (van de Water, Perenboom, and Boshuizen (1996)). Moreover, it is particularly important to distinguish gender differences in the progress of (healthy) life expectancy, as patterns of mortality and health are not exactly the same for males and females (see, for example, Van Oyen, Cox, Jagger, Cambois, Nusselder, Gilles, and Robine (2010) and Van Oyen, Nusselder, Jagger, Kolip, Cambois, and Robine (2013)).

Sullivan (1971) proposed a method to estimate healthy life expectancy. He developed the concept of combining health prevalence data with mortality data in a period life table to measure healthy life expectancy. To apply Sullivan's method, one needs data from cross-sectional surveys or population censuses. Such data set is typically easy to obtain. Due to the often large sample sizes of these health surveys, they generally produce highly reliable estimates of age-specific prevalence. In addition, Sullivan's method is straightforward to apply, having many extensions, see, for example, Molla, Wagener, and Madans (2001), Manton, Gu, and Lamb (2006b), and Imai and Soneji (2007a). A detailed manual for applying Sullivan's approach is available in Jagger, Cox, Le Roy, and the EHEMU team (2006).

In this paper, we study the future (healthy) life expectancy, employing the approach by Sullivan (1971), by taking into account the joint development of mortality, health, and two macroeconomic variables, namely real GDP per capita and the unemployment rate. Moreover, uncertainties of future developments in (healthy) life expectancy are quantified. In particular, the paper not only quantifies the uncertain development of mortality, like most of the other literature does, but also quantifies the joint uncertain evolutions of mortality and health, in combination with real GDP per capita and the unemployment rate. We consider macroeconomic variables in this study because those factors may help to predict the development of health and mortality, see

for instance, Van Oyen, Cox, Jagger, Cambois, Nusselder, Gilles, and Robine (2010), Thorslund, Wastesson, Agahi, Lagergren, and Parker (2013), Yang, De Waegenaere, and Melenberg (2013b), Niu and Melenberg (2014), and Liu, Arai, Kanda, Lee, Glasser, and Tamashiro (2013). In our study we use U.S. data from 1972 to 2010. We model mortality and health adopting the method introduced by Yang, De Waegenaere, and Melenberg (2013b), who include observed variables in the Lee and Carter (1992) framework to model health. We predict the joint future dynamics of the variables of interest using a vector autocorrelation (VAR) model.

Another method commonly adopted to derive healthy life expectancy is the multistate life table method, which is proposed by Schoen and Woodrow (1980). Differently from Sullivan's approach, a multistate life table method allows for recovery from certain health conditions (for example, from unhealthy to healthy). However, such a method requires a longitudinal data set, which is typically difficult to get. Furthermore, many researchers including Hayward and Grady (1990), Crimmins, Hayward, and Saito (1994), Crimmins, Hayward, Saito, and Kenkyūjo (1996), and Laditka (1998) pointed out that because the longitudinal data set is often collected every one, two, or even four to five years, health transitions may happen over relatively long time intervals only (for example, the National Long Term Care Survey in the United States). For these reasons, many researchers apply Sullivan's method. Comprehensive reviews on healthy life expectancy studies are provided by Laditka and Laditka (2002), Robine, Romieu, and Michel (2003), and Laditka and Laditka (2009).

In many studies, forecasts of healthy (or disability free) life expectancy allow changes in future mortality rates, but assume that future health remains constant or has different deterministic scenarios, see, for example, Jagger, Matthews, Spiers, Brayne, Comas-Herrera, Robinson, Lindesay, and Croft (2006), Jacobzone (2000), and Manton, Gu, and Lamb (2006a). A disadvantage of such deterministic forecast methods is that they do not provide any information regarding the likelihood of future health changes (Jagger, Cox, Le Roy, and the EHEMU team (2006) and Van Baal, Peters, Mackenbach, and Nusselder (2013)). Papers that apply a stochastic approach include Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012) and Van Baal, Peters, Mackenbach, and Nusselder (2013) who use Lee and Carter (1992) framework for the Dutch population and also provide corresponding prediction intervals. Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012) derive projections of disability-free life expectancy and Van Baal, Peters, Mackenbach, and Nusselder (2013) model differences in (healthy) life expectancy due to differences in educational levels. Our approach differs from these approaches in that we model the developments of aggregated health and mortality, in combination with important macroeconomic determinants.

As discussed by Yang, De Waegenaere, and Melenberg (2013b), extending the Lee and Carter (1992) by including observed macroeconomic quantities has some advantages. In particular, capturing the time variation by macroeconomic variables might

result in more precise model-based forecasts, if these macroeconomic variables are able to capture most of the time variation. Without such macroeconomic variables the trend in the Lee and Carter (1992) model is fully captured by a latent time variable. This might result in some overfitting in order to capture the trend as good as possible. The latent time variable will then be more volatile than the macroeconomic variables, potentially resulting in wider model-based forecast intervals. See also Niu and Melenberg (2014). In addition to these advantages, in our model the time variation in mortality and health are both captured to a large extent by the macroeconomic variables. This implies that we do not only capture the correlation between the time variation in mortality and health via the residuals, as in, for example, Majer, Stevens, Nusselder, Mackenbach, and van Baal (2012), but also directly via the common time varation in the macroeconomic variables. This will avoid deviating trends.

The remainder of this paper is organized as follows. The methodology applied in this paper will be introduced in the next section. The data used will be presented in Section 5.3. Next, Section 3.4 shows the empirical results of our study on (healthy) life expectancy, focusing on newborns and 65 years old using data from 1972 to 2010 and presenting forecasts to 2020 in the United States, distinguished by genders. Finally, the paper concludes in Section 5.5.

3.2 Methodology

This section illustrates the methodology we use to estimate and forecast life expectancy and healthy life expectancy.

3.2.1 Life Expectancy and Healthy Life Expectancy

Our starting point is the (raw) central death rate $m_{x,t}$ for a cohort born in year $t-x$ over a one year age interval $[x, x+1)$. Next, let $q_{x,t}$ be the one year death probability of age class x at time t (i.e., the time t conditional probability of death within an age interval of length one, given that an individual at age x survived up to age x. We can approximate $q_{x,t}$ on the basis of $m_{x,t}$ as (see Molla, Wagener, and Madans (2001)),

$$q_{x,t} = \frac{m_{x,t}}{1 + (1 - a_{x,t})m_{x,t}}, \tag{3.1}$$

where $a_{x,t}$ is the average proportion of years lived in the age interval $[x, x+1)$ among those who are alive at age x but die within 1 year. Using the one year death probabilities $q_{x,t}$, we define the τ year survival probabilities ${}_\tau p_{x,t}$ as

$${}_0 p_{x,t} = 1, \text{ and } {}_\tau p_{x,t} = \prod_{j=1}^{\tau} \left(1 - q_{x+j-1,t+j-1}\right), \tau \geq 1.$$

Life expectancy is then given by

$$LE_{x,t} = \sum_{\tau \geq 0} \left({}_{\tau+1}p_{x,t} + {}_{\tau}p_{x,t} q_{x+\tau,t+\tau} a_{x+\tau,t+\tau} \right). \tag{3.2}$$

If, in this equation, we set $q_{x+\tau,t+\tau} = q_{x+\tau,t}$, for all $\tau \geq 0$, then we use the one year death probabilities that can be obtained from the period life table at time t. If, on the other hand, each $q_{x+\tau,t+\tau}$ is the one year death probability applying to age class $x + \tau$ at year $t + \tau$, then the life expectancy is based on a cohort life table. In this paper, we calculate life expectancy on the basis of cohort life tables.

Next, let $\pi_{x,t} \in [0, 1]$ represent the health status index of age class x in year t, i.e., the fraction of the subpopulation of age class x that is in "poor health" in year t. The fraction in "good health" is then given by $1 - \pi_{x,t}$. In our analysis of healthy life expectancy we follow Imai and Soneji (2007a). The healthy life expectancy is then defined as

$$HLE_{x,t} = \sum_{\tau \geq 0} (1 - \pi_{x+\tau,t+\tau}) \times \left({}_{\tau+1}p_{x,t} + {}_{\tau}p_{x,t} q_{x+\tau,t+\tau} a_{x+\tau,t+\tau} \right). \tag{3.3}$$

Similar to life expectancy, we shall calculate healthy life expectancy on the basis of cohort life tables. Thus, each $\pi_{x+\tau,t+\tau}$ is the health status index applying to age class $x + \tau$ at year $t + \tau$.

In these equations $q_{x+\tau,t+\tau}$ and $\pi_{x+\tau,t+\tau}$, $\tau \geq 1$, refer to *future* one year death probabilities and *future* health status indices from the perspective of year t. We shall consider these quantities to be random at time t. As a consequence, both life expectancy $LE_{x,t}$ and healthy life expectancy $HLE_{x,t}$ will be random variables at year t. In the sequel, we shall present "best estimates" of these quantities at time t, with corresponding confidence intervals.

3.2.2 Modeling and jointly forecasting mortality and health

A cohort life table method requires a very long data period, a large part of which typically has to be forecasted. When determining life expectancy, existing studies have taken into account the uncertain development of future mortality. However, few studies concentrate on forecasting both life expectancy and healthy life expectancy, while accounting for the uncertainty in the health development and the uncertainty in the joint evolution of mortality and health. We apply a methodology that allows to incorporate the joint movement of mortality and health, based on Yang, De Waegenaere, and Melenberg (2013b), who study the development of health, also including the joint development with real GDP per capita and the unemployment rate. In this section, we first describe the model used to capture the development of mortality and health, then introduce a prediction approach which allows to capture the joint developments of mortality, health, GDP, and the unemployment rate.

In the mortality literature, the approach proposed by Lee and Carter (1992) is well documented and widely used. Recently, Yang, De Waegenaere, and Melenberg (2013b) applied the Lee-Carter framework to model the trend in health. They also extended this model by incorporating economic variables. They found that the macroeconomic variables included, namely real GDP per capita and the unemployment rate, are able to capture a large part of the health trend instead of the single latent variable in the original Lee-Carter model. In this paper, we adopt Yang, De Waegenaere, and Melenberg (2013b) to model the trends in both mortality and health.

To describe the Lee-Carter model (for either males or females) with observed variables included, let $y^c_{x,t}$ denote either the mortality rate $m_{x,t}$, if $c = m$, or the health prevalence ratio $\pi_{x,t}$, if $c = \pi$, at age x and year t of a cohort born in year $t - x$, and let Z_t be an m-dimensional vector containing the components of observed macroeconomic variables. In our case, $m = 2$ and the included macroeconomic variables are logarithm of real GDP per capita and the unemployment rate. In the Lee-Carter framework, we assume that some transformation F of $y^c_{x,t}$ satisfies the following decomposition,

$$F\left(y^c_{x,t}\right) = \alpha^c_x + \beta^c_x \kappa_t + (\rho^c_x)' Z_t + \epsilon^c_{x,t}. \tag{3.4}$$

In this decomposition, α^c_x describes the time-independent level of mortality or health as a function of age x, κ^c_t is a time-dependent univariate latent variable and represents the change in the overall level of $F\left(y^c_{x,t}\right)$ over time, β^c_x describes the age group-specific sensitivity of the overall changes when κ^c_t varies, $\rho^c_x = \left(\rho^c_{x,1}, \cdots, \rho^c_{x,m}\right)'$ is an m-dimensional age group-specific parameter vector, containing the coefficients corresponding to the components in Z_t, and $\epsilon^c_{x,t}$ is the error term, reflecting idiosyncratic time- and group-specific influences, with mean 0 and (possibly group-specific) variance $\sigma^2_{x,c}(\epsilon^c_{x,t})$.

As discussed by Yang, De Waegenaere, and Melenberg (2013b), certain constraints are needed in order to uniquely identify parameters β^c_x, κ^c_t, and ρ^c_x, namely

$$\sum_t \kappa^c_t = 0, \sum_x \beta^c_x = 1, \sum_x \rho^c_{x,i} = 1, \text{ for each i = 1, \ldots, m.} \tag{3.5}$$

The first constraint implies that for each x the estimate for α^c_x will be an average of the $F\left(y^c_{x,t}\right)$ over time. The second one implies that β^c_x represents which fraction (over all groups) of the change in κ^c_t is captured by group x.[1]

We choose the logarithm transformation as the functional form F for both the cen-

[1] As argued by Cairns, Blake, Dowd, Coughlan, Epstein, Ong, and Balevich (2007) and Pitacco, Denuit, Haberman, and Olivieri (2009a), the first constraint is a natural choice, whereas other choices of the second constraint have no impact on the quality of the fit, nor the model forecasts. Other constraints can be found in the literature, for instance, Wilmoth (1993) employs $\sum_x (\beta^c_x)^2 = 1$. For uniquely identifying ρ^c_x, Yang, De Waegenaere, and Melenberg (2013b) propose to include the third constraint and normalize the components of the vector Z_t such that they have mean 0 and variance 1.

tral mortality death rate ($m_{x,t}$), as Lee and Carter (1992) suggested, and for the health prevalence ratio ($\pi_{x,t}$), namely $F(y^c) = \log(y^c)$. Based on a sensitivity analysis Yang, De Waegenaere, and Melenberg (2013b) show that other investigated functional forms for $\pi_{x,t}$ do not provide better estimations and forecasts than the log-transformation. The Lee-Carter model with observed variables included can be estimated under the Newton-Raphson procedure, generalized by Renshaw and Haberman (2006). For details we refer to Yang, De Waegenaere, and Melenberg (2013b).

To predict mortality and health, we first need to predict all the time components associated with mortality and health in the described Lee-Carter framework with observed variables, namely κ^m, κ^π, and the observed variables. Including Lee and Carter (1992), most studies, see for instance Tuljapurkar, Li, and Boe (2000), use time series approaches to model and predict the future dynamics of κ_t^m. In this paper, we adopt the Vector Autocorrelation (VAR) model to fit and predict the interdependence of the four time components of interest, namely κ^m, κ^π, and the observed variables. Then, based on the forecasts of these independent variables from the VAR model, future values of the dependent variables m and π can be forecasted using the Lee-Carter model extended with the macroeconomic variables, see equation (3.4). Finally, we can construct a cohort life table from the observed and predicted mortalities. Moreover, (healthy) life expectancy can be derived following the process presented in section 3.2.1.

Because components in the VAR model need to be stationary, before applying a VAR model, we first have to test whether or not these variables, or their d^{th} order differences if these are going to be included, are stationary processes. Let K_t denote a vector with components κ_t^m and κ_t^π, then $D_d(K_t)$ and $D_d(Z_t)$ represent the d^{th} order difference of the vectors K_t and Z_t, respectively. The subscript d will be omitted when $d = 1$. If our four variables of interest are stationary processes after taking the d^{th} differences, a VAR(p) model can be written as

$$\boldsymbol{Y}_t \equiv \begin{bmatrix} D_d(K_t) \\ D_d(Z_t) \end{bmatrix} = \boldsymbol{C} + \sum_{i=1}^{p} \Theta_i \boldsymbol{Y}_{t-p} + \nu_t, \tag{3.6}$$

where p is the number of lags included, $\boldsymbol{C}$ in our case is a (4×1) parameter vector of constant, Θ_i in our case is a 4×4 coefficient matrix, and ν_t in our case is a 4-dimensional vector of white noise terms with means zero and covariance matrix Σ_ν. The lag length p can be determined by the three most common information criteria, namely the Akaike (AIC), Bayesian (BIC) and Hannan-Quinn (HQIC) information criteria. The values can be calculated as follows,

$$AIC(p) = -2LLF + 2N, \tag{3.7}$$

$$BIC(p) = -2LLF + 2N\log(T), \tag{3.8}$$

$$HQIC(p) = -2LLF + 2N\log(\log(T)), \tag{3.9}$$

where LLF is the value of the loglikelihood function with in our case $N = 4 + p4^2 + \frac{4(4+1)}{2}$ as the number of parameters. The process is to estimate VAR(p) models with orders $p = 0, \ldots, p_{max}$ accordingly, the value of p with the lowest AIC, BIC, or HQIC will be selected. A detailed discussion on the application of model selection criteria in VAR models is provided in chapter four in Lutkepohl (1991).

Once the parameters in the VAR(p) model are estimated, we are able to derive the point predictions for the four time series of interests s period ahead, conditional on $\boldsymbol{Y}_t$ at time t, that is $\widehat{\boldsymbol{Y}}_{t+s} = (D_d(\widehat{\boldsymbol{K}}_{t+s}), D_d(\widehat{\boldsymbol{Z}}_{t+s}))'$. In turn, the forecasts of $\log(\widehat{y}^c_{x,t+s})$ can be obtained from equation (3.4) with the estimated parameters $\widehat{\alpha}^c_x$, $\widehat{\beta}^c_x$, and $\widehat{\rho}^c_x$,

$$\widehat{y}^c_{x,t+s} = F^{-1}\left(\widehat{\alpha}^c_x + \widehat{\beta}^c_x \widehat{\kappa}^c_{x,t+s} + (\widehat{\rho}^c_x)' \widehat{Z}_{t+s}\right), c = m, \pi. \tag{3.10}$$

Cohort life tables can be constructed at this point, from which "best estimates" of (healthy) life expectancy (and corresponding confidence intervals) can be computed.

We construct confidence intervals around $LE_{x,t}$ and $HLE_{x,t}$, by taking into account two common uncertainties created from the VAR model. One is associated with randomness of v_t in equation (5.2), to which we shall refer as *process risk*. The other uncertainty is so-called *parameter risk*, which is present due to the inaccuracy of estimated model parameters in equation (5.2). In addition, we take into account *measurement risk*, i.e., the risk due to $\epsilon^c_{x,t}$ in (3.4) for both mortality ($c = m$) and health ($c = \pi$).[2].

3.3 Data description

In this section, we describe the data, namely the development of U.S. mortality, self-assessed health status, real GDP per capita, and the unemployment rate. For self-assesssed health and for the two macroeconomic variables, we use the same data as in Yang, De Waegenaere, and Melenberg (2013b). For completeness, we include a brief description of these data.

3.3.1 Mortality

This study obtains the consecutive annual cross-sectional mortality rate from 1972 to 2010 for the U.S. population from the Human Mortality Database (HMD)[3]. HMD is a collaborative project, involving research teams of the Department of Demography at the University of California, Berkeley, and of the Max Planck Institute for Demographic Research. It contains detailed annual age-specific mortality rates ($m_{x,t}$), which are defined as the ratios of the number of deaths to the total number of person-years

[2]Contrary to Yang, De Waegenaere, and Melenberg (2013b) we do not include parameter risk generated by the parameters in (3.4)

[3]The website of Human Mortality Database is http://www.mortality.org/.

in a population, for the one year age interval $[x, x+1)$ at time t, i.e.,

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}}. \tag{3.11}$$

where $D_{x,t}$ is the number of deaths and $E_{x,t}$ is the corresponding exposure (number of person-years).

In our study, the mortality rate consists of all males or all females in the U.S., at age x, with $x \in \{0, 1, \cdots, 110+\}$, where 110+ stands for the age category of age 110 and higher. Figure 3.1 shows the time pattern of the log central mortality rates for three age classes (age 25, 55, 85) in the examined period, for males (left panel) and females (right panel). The levels of the curves for females are lower than the corresponding curves for males. Figure 3.2 shows the percentage change in the log central mortality rates over the years 1972 to 2010 for all ages between 0 and 100 years. The relative change for males is higher (in absolute values) than the relative change for females for most age classes up to 90 years. Moreover, for males the relative change (in absolute values) is highest for the youngest and the elderly in the age range 70–90. For females the relative increase (in absolute value) is highest for the youngest and around age 90.

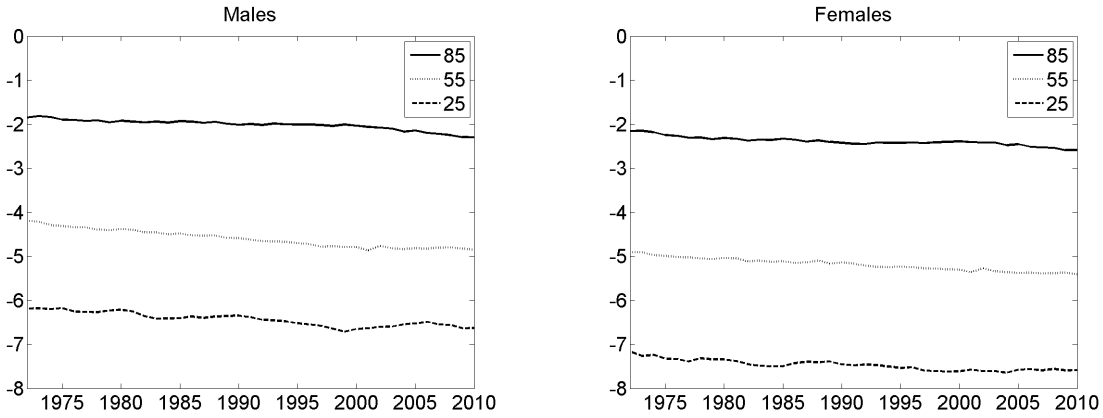

Figure 3.1 – Description of the central mortality rate in the U.S.

Data Source: Human Mortality Database (HMD), see `http://www.mortality.org/`
The log central mortality rate over time (1972–2010) for three ages: 25, 55, 85. Left panel: males; right panel: females.

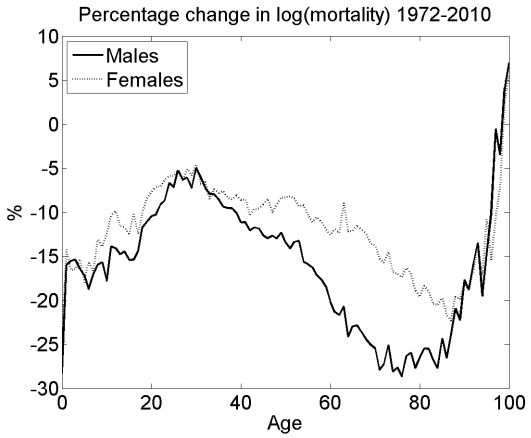

Figure 3.2 – Relative change in log central mortality rate in the U.S.

Data Source: Human Mortality Database (HMD), see `http://www.mortality.org/`
The relative change in log central mortality rate between 1972 and 2010, as function of age.

3.3.2 Self-reported health

Self-assessed health is a simple subjective measure of health that provides an ordinal ranking of perceived health status. As discussed in Yang, De Waegenaere, and Melenberg (2013b) it is a commonly used measure of health. Current research studied the relationship, for instance, between self-assessed health and social economic status, (see Smith (1999), Contoyannis, Jones, and Rice (2004), Robine, Romieu, and Michel (2003)), and the relationship between health and life styles (see Contoyannis and Jones (2004) and Crimmins, Hayward, and Saito (1994))

The National Health Interview Survey (NHIS) is the principal nationally-representative source of data on health of the civilian noninstitutionalized population in the United States. It provides the consecutive annual cross-sectional self-assessed health status of individuals from newborn to age class 85+ at the national level.[4] We use the annual data from 1972 until 2010. Throughout each year, survey respondents are asked to rate their health with the conditions of "excellent," "very good," "good," "fair," or "poor." In this paper, the health status index defines the fraction of the population with a particular health status ("good" or "bad") at a particular age in a certain year. Following Yang, De Waegenaere, and Melenberg (2013b), we choose to cluster "poor" and "fair" health status as "bad" health condition, and the three other categories as "good" health condition. [5]

[4]Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 4.0. Minneapolis: University of Minnesota, 2011. The health data were downloaded as the variable "Health status" available from the website `https://www.ihis.us/ihis-action/variables/group/health_general`.

[5]See Yang, De Waegenaere, and Melenberg (2013b) for further motivation. Moreover, although such an aggregation method may remove some potential information, it avoids the problem caused by the

Let $\pi_{x,t}$ denote the health status index for a group with age x at time t with a particular health condition, adjusted for the inverse probability of a person being selected into the sample (i.e., the person weight). It is computed as follows:

$$\hat{\pi}_{x,t} = \frac{1}{\sum_{j=1}^{N_{x,t}} w_{j,x,t}} \sum_{j=1}^{N_{x,t}} w_{j,x,t} H_{j,x,t}, \tag{3.12}$$

where $N_{x,t}$ represents the total number of survey respondents of group x, $H_{j,x,t}$ is a zero-one indicator of particular health condition for the j^{th} respondent in this group, and $w_{j,x,t}$ is the person weight. A more detailed description of the self-assessed health data, including response rates and person weight can be found in Yang, De Waegenaere, and Melenberg (2013b), section 3.1. In this study, we define $H_{j,x,t} = 1$ when respondent j reports a "bad" health condition ("poor" or "fair"), and $H_{j,x,t} = 0$ otherwise. Therefore, the health status index representing the fraction of the population who reports a particular condition of health, reflects the overall health level of the age-specific population in a given year. In our application, the population groups consist of all males or all females in the U.S., and a group x consists of all individuals in this population at age x, with $x \in \{0, 1, \cdots, 84, 85\}$, where 85 stands for the age category 85+.

Figure 3.3 describes the health status index averaged over age ($\bar{\pi}_{.,t}$, see the left panel) and averaged over time ($\bar{\pi}_{x,.}$, see the right panel), distinguished by genders. Moreover, Figure 3.4 presents the percentage change in the log health status index over the years 1972 to 2010 for all ages (from age 0 to age class 85+ years). As expected, $\bar{\pi}_{.,x}$, which represents people's "bad" health, has upward trends over age. This indicates that people's health condition is getting worse as people age in general. Over time, health trends are more volatile than mortality as shown in Figure 3.1. However, the relative increase (in absolute values) seems to be more evenly spread over the various age classes (see Figure 3.4). On average, females report a worse health status than males. This difference seem to be due to the difference between around age 15 and age 55.

small numbers found in the worst category, even in the highest age groups (Bebbington and Shapiro (2006)).

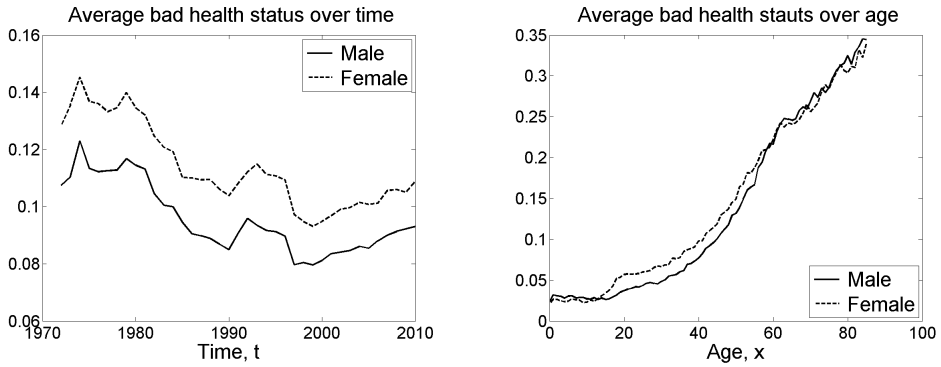

Figure 3.3 – Description of the Health Status Index in the U.S.

Data Source: Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 4.0. Minneapolis: University of Minnesota, 2011.
The fraction in the population, who are in bad health condition averaged over age ($\bar{\pi}_{.,t}$, see the left panel) is calculated based on the total number of respondents at each age group x.
The fraction in the population, who are in bad health condition averaged over time ($\bar{\pi}_{x,.}$, see the right panel) is calculated based on the total number of respondents in each survey year.

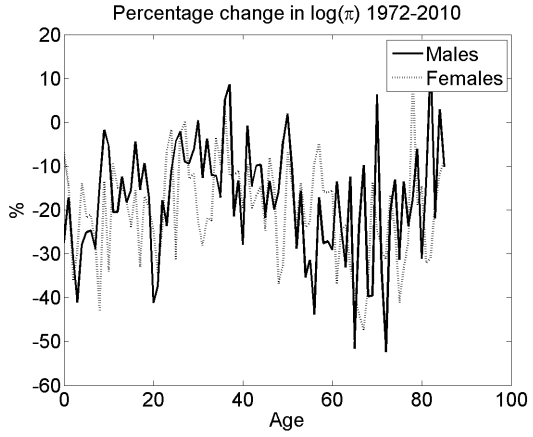

Figure 3.4 – Relative change in log health status index in the U.S.

Data Source: Minnesota Population Center and State Health Access Data Assistance Center, Integrated Health Interview Series: Version 4.0. Minneapolis: University of Minnesota, 2011.
The relative change in log central health status index between 1972 and 2010, as function of age.

3.3.3 Observed variables

As in Yang, De Waegenaere, and Melenberg (2013b), economic variables are taken into account when modeling health and mortality changes. We employ two macroeco-

nomic variables, namely gross domestic product per capita (GDP) corrected by inflation and the total unemployment rate in the United States. These are obtained from the Organization for Economic Cooperation and Development (OECD) Statistics Extracts (the Country Statistical Profiles, 2010).

Figure 3.5 describes the in sample evolution from 1972 to 2010 of the two macroeconomic variables, with the left panel for GDP per capita and the right panel for the unemployment rate (in %). GDP per capita, measured in U.S. dollar at constant prices (with OECD base year 2005), is trending upward over the examined years. OECD Statistics defines the total unemployment rate as the ratio of people out of work and actively seeking it to the population of working age either in work or actively seeking it. In the United States, the minimum age for employment set by the Fair Labor Standards Act (FLSA) is 14 years old. The OECD, therefore, provides the total unemployment rate for the U.S. working age population aged 15 years and over, all raw and seasonally adjusted. The right panel of Figure 3.5 shows the in-sample development of the total unemployment rate from 1972 to 2010 with an average 6.32% over years, the minimum 4% in 2000, and the maximum 9.7% in 1982.

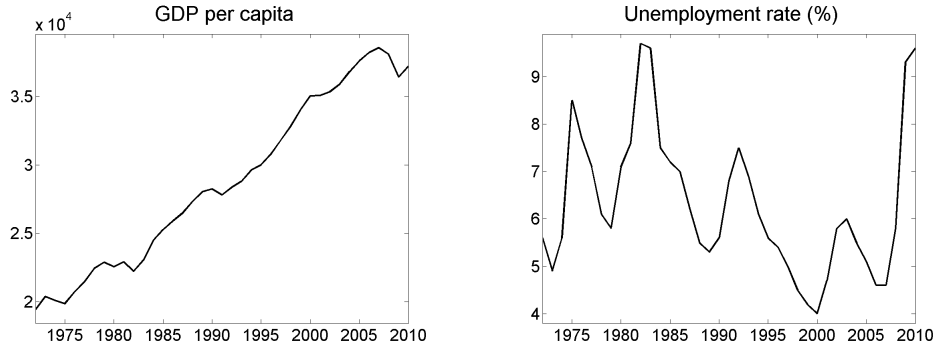

Figure 3.5 – Description of macroeconomic variables.

Note: Data Source: OECD.Stat Extracts, see `http://stats.oecd.org/`
The left graph describes the GDP per capita in US dollars in base year 2005. The right graph describes the total unemployment rate for working age group 15 years and older.

3.4 Empirical Results

This section presents the empirical results in terms of the remaining (healthy) life expectancy for the newborns and 65 year olds from 1972 to 2020. We first present the estimates related to modeling mortality and health, based on which future mortality and health are predicted. We then show the developments of life expectancy and healthy life expectancy for males and females from several aspects. In the next section, we also compare the results from the presented model with other alternative models, which either do not take into account the macroeconomic variables, or do not model

time components jointly.

3.4.1 Estimation results for mortality and health

Following the methodology described in section 3.2.2, we estimate both the Lee-Carter model with observed variables for mortality and for health, including real GDP per capita and the unemployment rate as observed variables, and using in both cases for F the log-transformation. Figure 4.8 and Figure 4.9 (see Appendix) present the estimates for male and female mortality and health respectively. We show the original estimates of the parameters, as well as their smoothed variants. Following Yang, De Waegenaere, and Melenberg (2013b), we present the transformed estimates of κ_t^c (and ρ_x^c), $c = m, \pi$, such that the transformed κ_t^c is orthogonal to Z_t, the observed variables the logarithm of real GDP per capita and the unemployment rate. See equations (7)–(9) in Yang, De Waegenaere, and Melenberg (2013b). The transformed κ_t^c for both $c = m$ and $c = \pi$ do not show a clear trend anymore, suggesting that real GDP per capita and the unemployment rate are able to capture most (if not all) of the trend in mortality and health. The coefficients ρ_x^c of the log real GDP per capita are negative for most age classes for both $c = m$ and $c = \pi$, meaning that an increase in real GDP per capita has a positive correlation with "survival" and "good health." The coefficients ρ_x^c of unemployment rate are also negative for most age classes for mortality ($c = m$), but the smoothed ρ_x^c of unemployment rate for health ($c = \pi$) are negative for the very young and very old, and positive other age classes (in terms of signs). This means that an increase in unemployment has a positive correlation with "survival", a positive correlation with "good health" for the very young and very old, but a negative correlation with "good health" for people at other age groups.

The (non-transformed) time variables κ_t^m for mortality and κ_t^π for health are combined with log real GDP per capita and the unemployment rate in a Vector Autoregression (VAR) to model and forecast their joint developments. First, as indicated by the Augmented Dickey-Fuller (ADF) tests, κ_t^m, κ_t^π, and log real GDP per capita are $I(1)$ processes. We therefore choose the first differences ($d = 1$) of these three variables to achieve a stationary VAR(p) process in equation (5.2). We also include the difference (instead of the level) of the unemployment rate. Thus, we consider $D(K_t) = K_t - K_{t-1}$ and $D(Z_t) \equiv Z_t - Z_{t-1}$. To determine p, models with $p = 0$ up to and including $p_{max} = 4$ are estimated and their values in terms of three information criteria are compared. Our results show that for both genders, both AIC and HQIC suggest $p = 3$, whereas BIC suggests $p = 0$. In the sequel, we consider a VAR(3) specification. A VAR(3) for the four variables of interest is written as follows,

$$\boldsymbol{Y}_t \equiv \begin{pmatrix} \Delta K_t \\ \Delta Z_t \end{pmatrix} = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \Theta_2 \boldsymbol{Y}_{t-2} + \Theta_3 \boldsymbol{Y}_{t-3} + \nu_t, \tag{3.13}$$

Table 3.1 and Table 4.1 (see Appendix) present the estimates of the VAR(3) models for males and females. We further use these estimates to forecast in the next section.

3.4.2 "Best estimates" life expectancy and healthy life expectancy

To compute the "best estimates" of life expectancy and healthy life expectancy until 2020, we need to construct "best estimates" of the cohort life tables with the first one starting in the earliest sample year 1972 and the final one starting in 2020. The data is available for mortality and health from 1972 to 2010, therefore, we need to employ forecasts from 2011 onwards.

Following the method described in section 3.2, we first forecast κ^m, κ^h, log real GDP per capita, and the unemployment rate based on the estimated VAR(3) model, see equation (4.12). These values can then be adopted as independent variables in equation (3.4) to forecast future mortality and health, see equation (3.10). The estimates $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\rho}_x$ in equation (3.10) are smoothed in order to avoid an erratic progression when predicting. In addition, when forecasting the mortality rate, we correct for the jump off bias. The one year death probabilities $q_{x,t}$ are calculated from observed ($t \leq 2010$) and forecasted ($t \geq 2011$) mortality rates, see equation (3.1). Here, we assume that $a_{x,t} = 0.5$ for cohorts at all single-year ages except at birth and at age 110. For age 110 (an open age interval) we set $a_{110,t} = \frac{1}{m_{110,t}}$ and $q_{110,t} = 1$. For infants, following the calculation of a_0 in Coale, Demeny, and Vaughan (1983) and Preston, Heuveline, and Guillot (2001), we adopt the following formulas. For males,

$$\begin{aligned} a_{0,t} &= 0.35, \text{ if } m_{0,t} \geq 0.107. \\ a_{0,t} &= 0.045 + 2.684 \times m_{0,t}, \text{ if } m_{0,t} < 0.107. \end{aligned} \tag{3.14}$$

And for females,

$$\begin{aligned} a_{0,t} &= 0.33, \text{ if } m_{0,t} \geq 0.107. \\ a_{0,t} &= 0.053 + 2.800 \times m_{0,t}, \text{ if } m_{0,t} < 0.107. \end{aligned} \tag{3.15}$$

Once $q_{x,t}$ is calculated for all x and t (including the relevant future time periods), we can construct a cohort life table, from which the "best estimates' of life expectancy can be calculated, using (3.2).

Next, we compute the "best estimates' of healthy life expectancy for each cohort by incorporating the (observed or forecasted) health status index into the cohort life table. we follow a similar procedure as described above for mortality. A major difference is that the oldest age group in the observed health sample data is the group 85+. We do have data on separate ages larger than 85, but only for the years 1972–1995. In Figure 3.6 we plot the 10%, 50%, and 90% quantiles of $\pi_{x,t}$ for both males and females and all

ages $x \in \{0, \cdots, 99+\}$, using only $t = 1972, \cdots, 1995$. We see that after age 85, there is no clear trend anymore, but, instead, quite a lot of uncertainty. We shall assume that $\pi_{x,t} = \pi_{85+,t}$, for $x = 85, 86, \cdots, 110$. In addition, when quantifying the corresponding confidence intervals, we shall explicitly take into account the large variation in $\pi_{x,t}$ for high ages. We do this by incorporating measurement risk, i.e., the risk due to the measurement errors $\epsilon^c_{x,t}$, see equation (3.4). For further details, please refer to the Appendix.

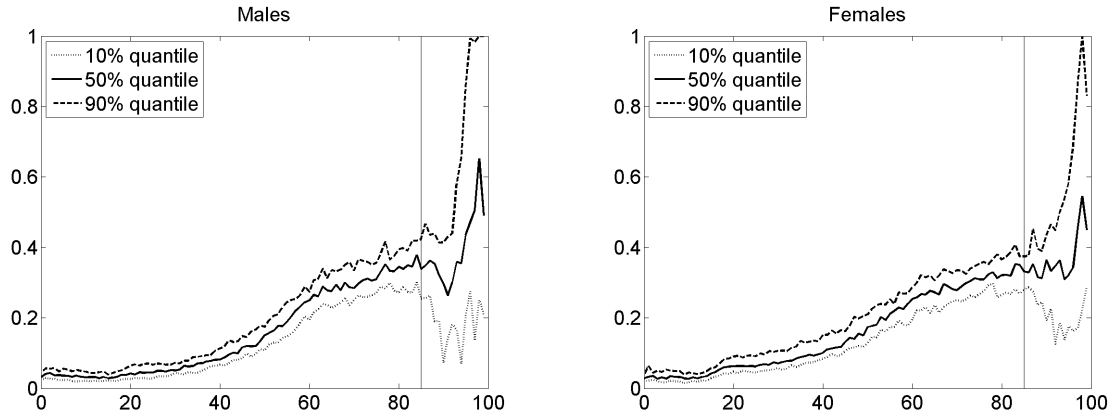

Figure 3.6 – Quantiles (10%, 50%, 90%) of $\pi_{x,t}, x = 0, \cdots, 99+$ for males and females for the subperiod 1972–1995.

Figures 3.7 and 3.8 show the comparison of "best estimates" of life expectancy (left panel) and healthy life expectancy (right panel) by genders for cohorts at birth and age 65, respectively. The best estimate results imply that males' life expectancy and healthy life expectancy are lower but increasing faster than females', both at birth and age 65. To illustrate, when we compare gender gaps in (healthy) life expectancy for cohorts at birth in 1972 and 2020, the gap in life expectancy reduces from 3.83 to 0.87 years (i.e. by 77.2%), and from 2.62 to 0.85 years (i.e., by 67.5%) in healthy life expectancy. The gender gaps for cohorts at age 65 are larger than at birth, but they are also reducing: The gap in life expectancy reduces from 4.56 to 1.81 years (60.4%), and reduces from 3.02 to 1.50 years (50.4%) in healthy life expectancy.

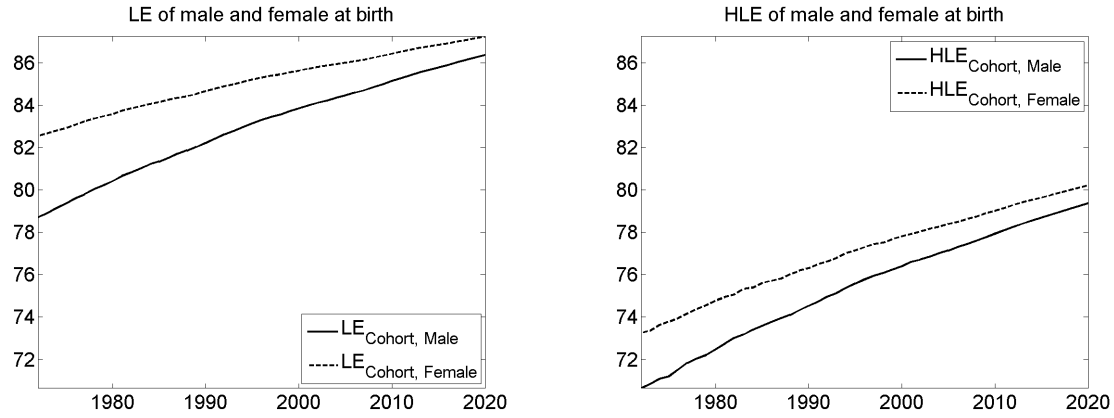

Figure 3.7 – Life expectancy and healthy life expectancy for males and females at birth

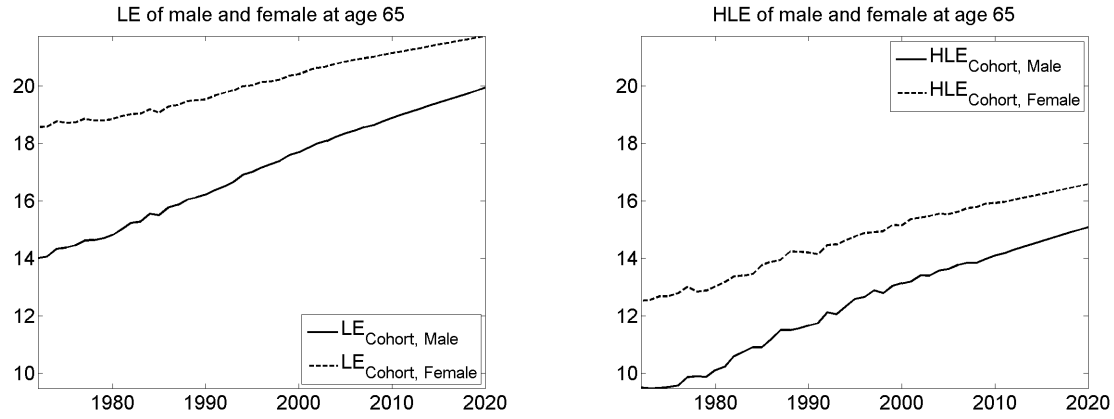

Figure 3.8 – Life expectancy and healthy life expectancy for males and females at age 65

To further understand how much (healthy) life expectancy increases, we compute values of relative increases in (healthy) life expectancy compared with the ones in the first available sample period. Let t_1 denote the (first sample) year 1972, the relative increases in life expectancy ($RLE_{x,t}$) and healthy life expectancy ($RHLE_{x,t}$) describe how much the (healthy) life expectancy of cohorts at age x in year t increases compared with cohorts at the same age x, but in year t_1. They are calculated as,

$$RLE_{x,t} = \frac{LE_{x,t}}{LE_{x,t_1}}, \quad RHLE_{x,t} = \frac{HLE_{x,t}}{HLE_{x,t_1}}.$$

Figure 3.9 presents the relative increases of (healthy) life expectancy of cohorts at birth and at age 65 in each year relative to the ones in 1972, distinguished by genders. Figure 3.9 shows that for both males and females, healthy life expectancy of cohorts at

birth (left panel) and age 65 (right panel) increase faster than life expectancy. Our best estimates results are in line with findings from Nusselder, Looman, Oyen, Robine, and Jagger (2010) and Van Oyen, Nusselder, Jagger, Kolip, Cambois, and Robine (2013). On the one hand, females' mortality advantage leads to their higher (healthy) life expectancy compared with males; On the other hand, females' larger prevalence of bad health condition reduces this gender gap in the future both at birth and age 65.

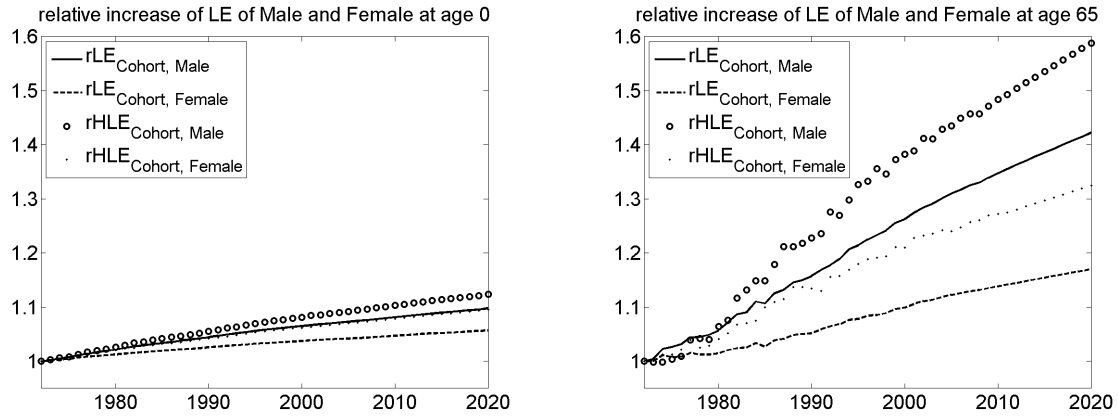

Figure 3.9 – Relative increases of life expectancy and healthy life expectancy for males and females at birth and age 65

3.4.3 Confidence intervals

So far, the discussed results are all based on the "best estimates" without considering any uncertainties. In this section, we quantify the uncertain development of (healthy) life expectancy by means of confidence intervals. These confidence intervals are calculated by taking into account two common types of risks when forecasting from the VAR model, namely process risk and parameter risk, see Section 3.2. In addition, we incorporate measurement risk, due to the measurement errors in equation (3.4).

Figure 3.10 shows the 95% age-specific confidence intervals of life expectancy and healthy life expectancy at birth by gender. Figure 3.11 presents these values for the cohorts at age 65. In both figures, the narrower confidence intervals are the 95% confidence intervals when only taking into account process and measurement risk, and the wider confidence intervals are the ones when considering next to process and measurement risk also parameter risk. In line with the findings in Van Baal, Peters, Mackenbach, and Nusselder (2013) for the Netherlands, our results for the United States suggest that life expectancy of cohorts at birth and age 65 are very likely to continue rising in the future. This also applies to healthy life expectancy at age 65 for both males and females. However, the lowest 95% confidence band for healthy life expectancy at birth reflects some uncertainty about a continuing increasing trend in these healthy life expectancies.

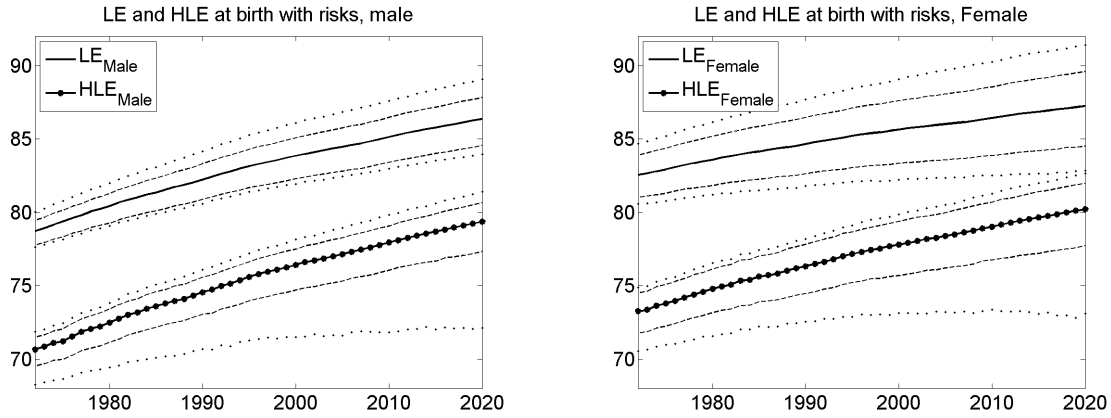

Figure 3.10 – Quantified uncertainties of (healthy)life expectancy of cohorts at birth for males and females

Note: Two 95% confidence intervals are presented in each graph. The narrower dashed curves present the 95% confidence interval when considering the process risk in the VAR model and measurement risk in the Lee-Carter model with observed variables. The wider dotted curves present the 95% confidence interval when considering both process and parameter risks in the VAR model, and measurement risk in the Lee-Carter model with observed variables.

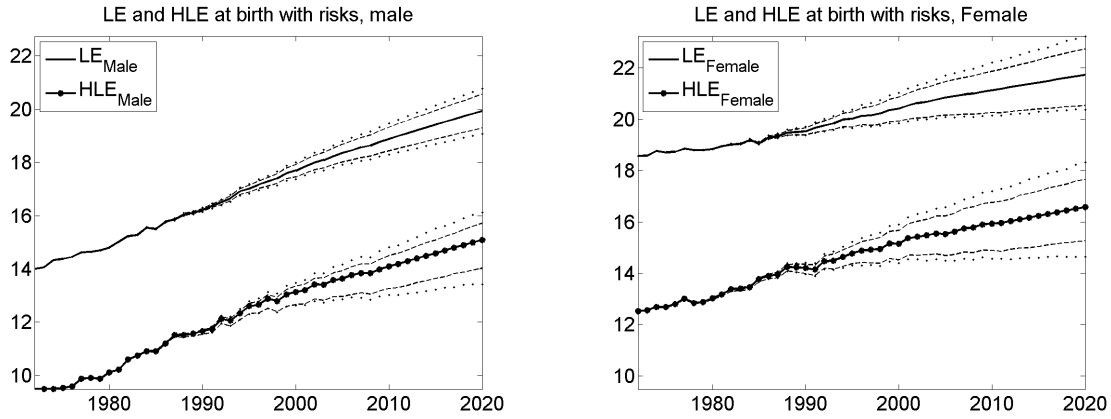

Figure 3.11 – Quantified uncertainties of (healthy)life expectancy of cohorts at age 65 for males and females

Note: Two 95% confidence intervals are presented in each graph. The narrower dashed curves present the 95% confidence interval when considering the process risk in the VAR model and measurement risk in the Lee-Carter model with observed variables. The wider dotted curves present the 95% confidence interval when considering both process and parameter risks in the VAR model, and measurement risk in the Lee-Carter model with observed variables.

Figure 3.12 presents the development of the proportion of healthy life expectancy over the total expected life years (HLE/LE) at birth and at age 65 by gender. The

corresponding 95% confidence intervals are based on process risk and measurement risk. The graphs show that the ratios for males and females for the cohorts at birth seem to keep rising and at the same time converging to each other. The ratios for the cohorts at age 65 also seem to keep rising, but the ratio of females seems to be slightly higher than the ratio of males, although this difference is not significant.

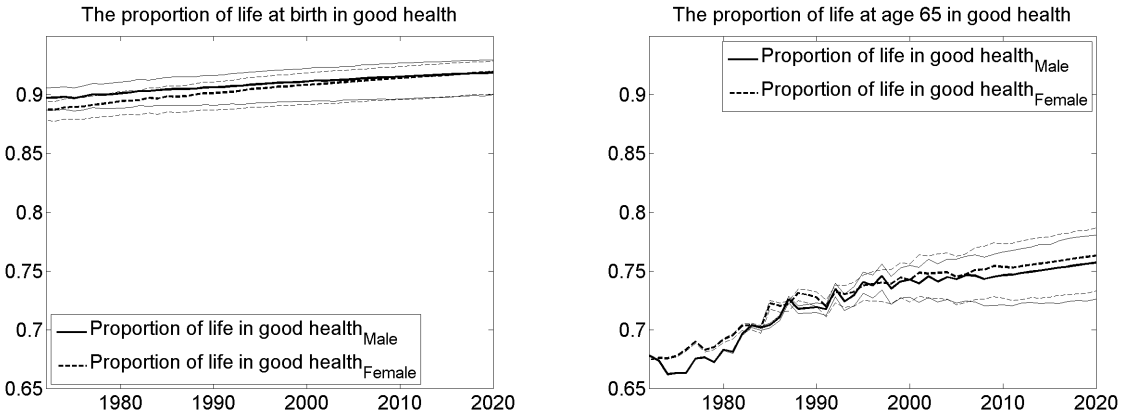

Figure 3.12 – The percentage of life at birth lived in good health

Note: The left panel shows the proportion of remaining lifetime at birth by genders in good health. The right panel shows the proportion of remaining lifetime at age 65 in good health. 95% confidence intervals generated from the process risks are presented.

3.4.4 Comparison with other models

In this subsection, we compare our findings with three alternative approaches, which do not jointly model mortality and health, do not consider the macroeconomic variables, or both. We shall refer to the model investigated so far as model "JoLCO" (with "Jo" standing for modeling the joint development of mortality and health, "LC" standing for Lee-Carter, and "O" for using observed macroeconomic variables). The first alternative method we consider is the Lee-Carter model with macroeconomic variables, but separately modeling the development of mortality and health. This variant is denoted as "SeLCO" (with "Se" standing for modeling the development of mortality and health separately). The second alternative models mortality and health by the original Lee-Carter model, without observed variables, but with joint modeling the development of mortality and health latent trends, to be denoted by "JoLC". The last one models mortality and health by the original Lee-Carter model, with separately modeling the development of mortality and health, to be denoted by "SeLC". In case of model JoLC a VAR(4) model is selected, following the same procedure as for model JoLCO. In case of models SeLC and SeLCO, we model and forecast mortality and health separately using an autoregression of order 4 ($AR(4)$).

A comparison of the four models only using the "best estimates" is provided in Figure 3.13 and Figure 3.14.

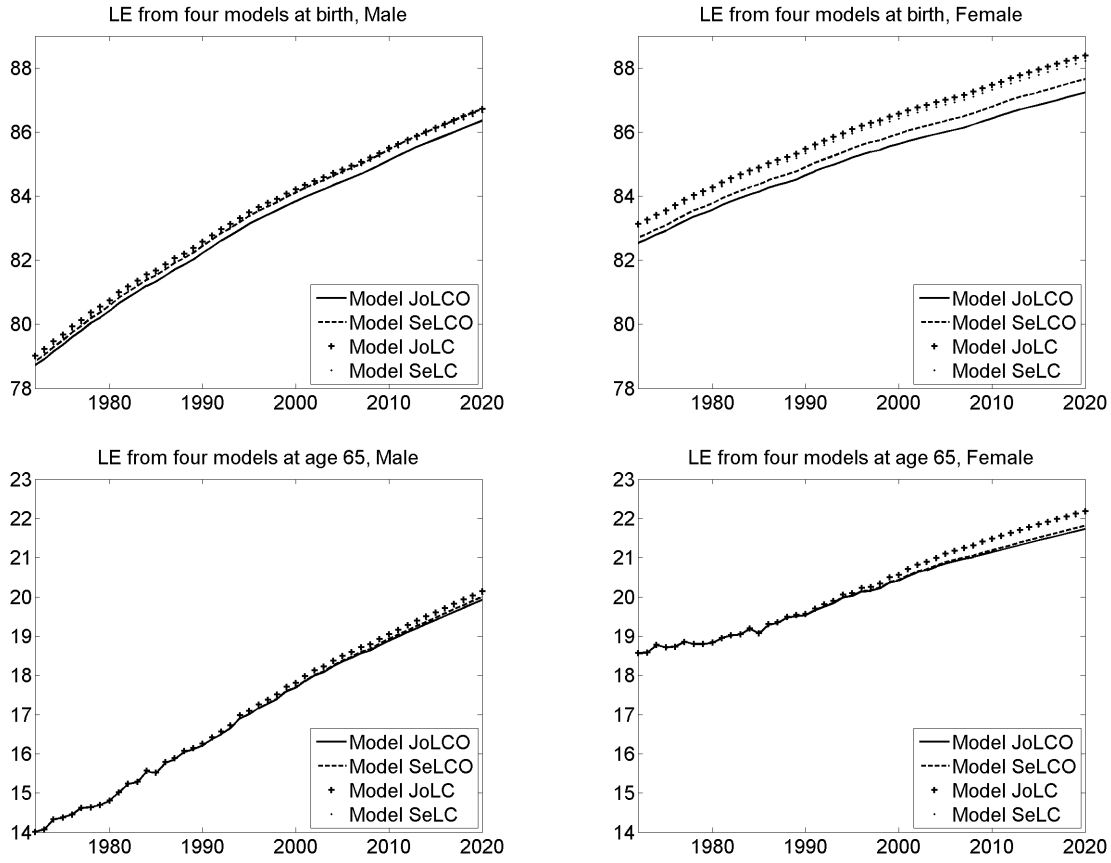

Figure 3.13 – Comparison of life expectancy from the four models of interest

Note: Graphs show life expectancy of cohorts at birth (the upper panel) and at age 65 (the lower panel) from four compared models.
'Model JoLCO' models mortality and health by Lee-Carter model with macroeconomic variables, and jointly models and forecasts latent trends of mortality and health in a VAR(4) model;
"Model SeLCO" models mortality and health by Lee-Carter model with macroeconomic variables, but separately models and forecasts latent trends of mortality and health in a AR model;
'Model JoLC' models mortality and health by Lee-Carter model, and jointly models and forecasts latent trends of mortality and health in a VAR(4) model;
"Model SeLC" models mortality and health by Lee-Carter model, but separately models and forecasts latent trends of mortality and health in a AR model.

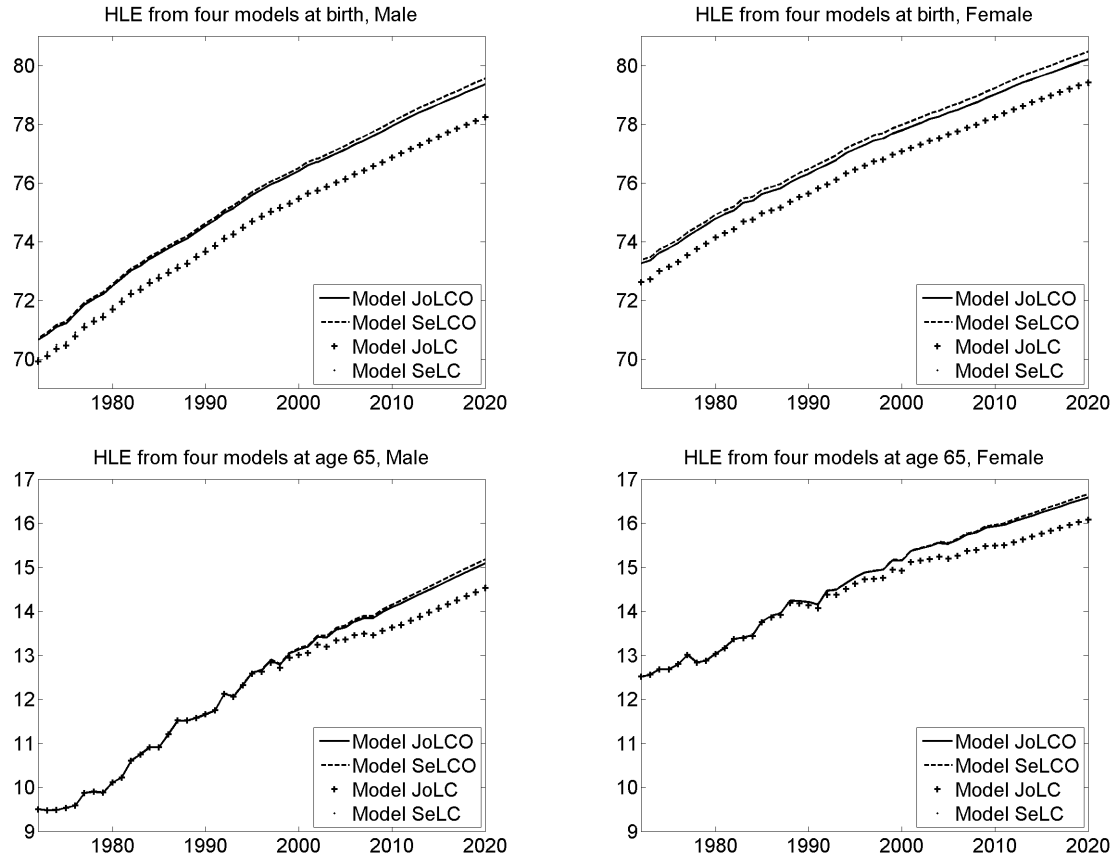

Figure 3.14 – Comparison of healthy life expectancy from the four models of interest

Note: Graphs show life expectancy of cohorts at birth (the upper panel) and at age 65 (the lower panel) from four compared models.
'Model JoLCO' models mortality and health by Lee-Carter model with macroeconomic variables, and jointly models and forecasts latent trends of mortality and health in a VAR(4) model;
"Model SeLCO" models mortality and health by Lee-Carter model with macroeconomic variables, but separately models and forecasts latent trends of mortality and health in a AR model;
'Model JoLC' models mortality and health by Lee-Carter model, and jointly models and forecasts latent trends of mortality and health in a VAR(4) model;
"Model SeLC" models mortality and health by Lee-Carter model, but separately models and forecasts latent trends of mortality and health in a AR model.

Figures 3.15 and 3.16 show the comparisons of life expectancy (the upper panel), and healthy life expectancy (the lower panel) with their 95% confidence intervals (representing process risk and measurement risk) of model JoLCO and the model that differs most from model JoLCO, i.e., model JoLC in case of life expectancy and model SeLC in case of healthy life expectancy. For males, we only find small differences in case of life expectancy, with a substantial overlap in terms of the forecast intervals. However, in case of healthy life expectancy we see some substantial differences, even though the

forecast intervals still overlap. Compared to model JoLCO that takes the joint development of mortality and health into account and that incorporates the development in observed economic variables, model SeLC, that ignores the joint development and that does not incorporate the development in observed economic variables, gives much wider forecast intervals for the newborns and seems to underestimate the future development for males at age 65. The wider forecast intervals in case of model SeLC might be due to the following reason, see also Niu and Melenberg (2014). Model SeLC only uses latent variables to capture the trend in mortality and health. This might result in some overfitting of the trend, generating extra volatility in the latent variable. This extra volatility will then be translated into a wider forecast interval.

In case of females, we also see some differences when comparing life expectancy at birth. Model JoLC that does not take into account the development in observed economic variables seems to overestimate the future life expectancy at birth compared to model JoLCO, that incorporates the development in observed economic variables. For healthy life expectancy we find similar differences for females as in case of males when comparing model SeLC and model JoLCO: model SeLC results in wider forecast intervals for the newborns and seems to underestimate the future development for females at age 65.

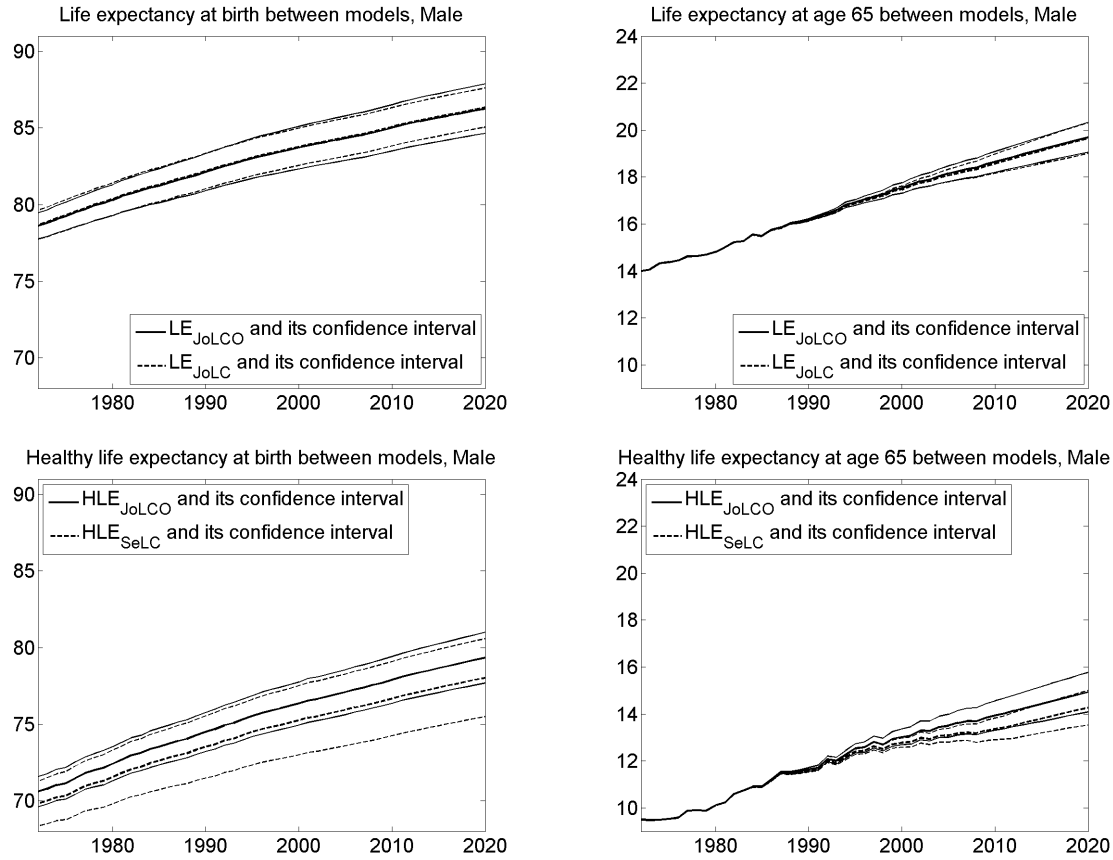

Figure 3.15 – Life expectancy and healthy life expectancy comparison for males, Model "JoLCO" versus model "JoLC" and versus model "SeLC."

Note: The upper panel shows life expectancy (left) and healthy life expectancy (right) of cohorts at birth with process risk and parameter risk. The lower panel shows life expectancy (left) and healthy life expectancy (right) of cohorts at age 65 with process risk and parameter risk.

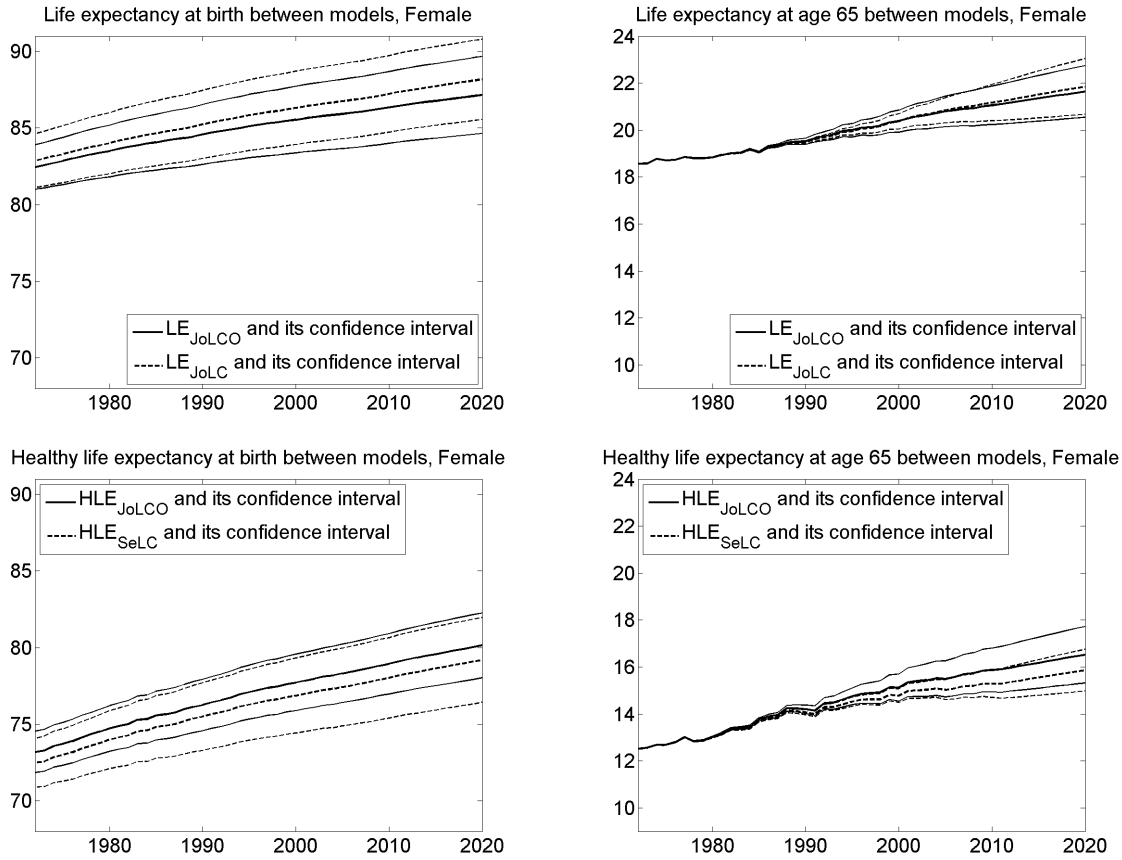

Figure 3.16 – Life expectancy and healthy life expectancy comparison for females, model 'JoLCO" versus model "JoLC" and versus model "SeLC."

Note: The upper panel shows life expectancy (left) and healthy life expectancy (right) of cohorts at birth with process risk and parameter risk. The lower panel shows life expectancy (left) and healthy life expectancy (right) of cohorts at age 65 with process risk and parameter risk.

3.5 Conclusion

This paper uses the data in the United States on the age-specific prevalence rates of self-assessed health from the National Health Interview Survey and the mortality prevalence from Human Mortality Database from 1972 to 2010 to jointly model and predict time trends in mortality, health, and macroeconomic variables. We compute the life expectancy and healthy life expectancy from cohort life tables using Sullivan's method. In addition to the point prediction, we also quantify two corresponding uncertainties, namely "process risk" and "parameter risk." Furthermore, measurement risk in the Lee-Carter framework is taken into account as well.

Our findings suggest that, between 1972 and 2010, there has been an improvement in the quality of life at age 65 in both absolute terms (number of years) and relative

terms (the proportion of life). Our point forecasts up to 2020 show a further increase. However, when also taking into account the forecast uncertainty, our forecast intervals become so wide that a decrease in (healthy) life expectancy cannot be excluded.

We also compare our model with alternative models that do not include observed macroeconomic variables or that do not take the joint development in mortality and health into account. These are the standard Lee and Carter (1992) models, with or without taking the joint development in mortality and health into account. Our results show that our approach might help to get more precise model-based forecasts. The included macroeconomic variables seem to capture the time variation in mortality and health in a smoother way than happens in the standard Lee and Carter (1992) model.

In this paper we considered males and females separately. While studying the association between the increase in life expectancy and the changes in the prevailing patterns of health, researchers realize that understanding the gender disparities is very important for improving efficiency of government policies, such as allocation of health care resources. Two similar patterns are found in many aging countries. First, both females and males show rapid increases in life expectancy. However, many researchers, including Waldron (1993), Trovato and Heyen (2006) Glei and Horiuchi (2007), and Thorslund, Wastesson, Agahi, Lagergren, and Parker (2013), have shown that the female advantage in life expectancy at older age has been narrowing since the 1970s in the U.S. Second, in a country like the United States, although females at most ages still have a remarkable lower mortality than males, they tend to report a worse health status than males (Case and Paxson (2005)). This is the so called male-female health-survival paradox. An interesting topic of future research would be to extend our model to also take into account the joint development in male and female mortality and health. This would allow to test whether the gender gap has narrowed, taking into account the uncertainty in the development of mortality and health.

Appendix

We quantify the accuracy of the estimates in the VAR model as follows,

1. We estimate the VAR model (4.12), using $\boldsymbol{Y}_t = (\Delta K_t, \Delta Z_t)'$, where $K_t = (\widehat{\kappa}_t^m, \widehat{\kappa}_t^h)'$, and obtain the estimates $\hat{\boldsymbol{C}}$ and $\hat{\Theta}$. Residuals $\hat{v}_t$ are computed as $\hat{v}_t = \boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_t$, where $\hat{\boldsymbol{Y}}_t = (\hat{\boldsymbol{C}} + \hat{\Theta}\boldsymbol{Y}_{t-1})$, and its variance covariance matrix is derived as $\widehat{\Sigma}_v = \widehat{Cov}(\hat{v}_t)$.

2. To ensure that the bootstrapped residuals $\hat{v}_t^b$ have the same variance covariance matrix $\widehat{\Sigma}_v$, we first decompose $\widehat{\Sigma}_v = \boldsymbol{ADA}'$, where $\boldsymbol{D}$ is a unique diagonal matrix with positive entries along the principal diagonal and $\boldsymbol{A}$ is a unique lower triangular matrix with 1-s along the principal diagonal. We then construct $\hat{u}_t = \boldsymbol{A}^{-1}\hat{v}_t$.

Next, $\hat{u}_t^b$ are generated by sampling with replacement of $\hat{u}_t$. The bootstrapped residuals are in turn computed as $\hat{v}_t^b = \boldsymbol{A}\hat{u}_t^b$.

3. The bootstrapped series $\boldsymbol{Y}_t^b$ are obtained by $\boldsymbol{Y}_t^b = \hat{\boldsymbol{Y}}_t + \hat{v}_t^b$. This yields the computation of the estimated parameters $\hat{\boldsymbol{C}}^b$ and $\hat{\Theta}^b$.

4. Then, we generate a projection $\widehat{\boldsymbol{Y}}_{t+s}^b$ based on bootstrapped estimates $\hat{\boldsymbol{C}}^b$ and $\hat{\Theta}^b$. When quantifying the process risk in the VAR model together, we use the future errors $\hat{v}_{t+s}^b$, which are the 2000 simulated innovations in the process risk.

5. With the estimates of α_x^c, β_x^c, ρ_x^c, projected κ_{t+h}^c, and $\hat{Z}_{t+h}^b$, we can generate projections of $y_{x,t+h}^c$ for both males and females, and for $c = m$ and $c = \pi$. The future errors $\epsilon_{x,t+h}^c$ are sampled from the estimated $\epsilon_{x,t}^c$. The future errors $\epsilon_{x,t+h}^\pi$ for $x = 86, 87, \cdots$ are drawn from the estimated $\epsilon_{85,t}^\pi$, inflated by a factor in line with the increase in the corresponding standard deviations of $\pi_{x,t+h}$. For males we use as inflation factor $106.5030 - 2.6167x + 1.6150x^2/100$, and for females we use $267.7101 - 6.2266x + 3.6344x^2/100$, for $x = 86, 87, \cdots$.

Estimates of the Lee-Carter model with observed variables. Figure 4.8 show estimates for male and female mortality. Figure 4.9 show estimates for male and female health.

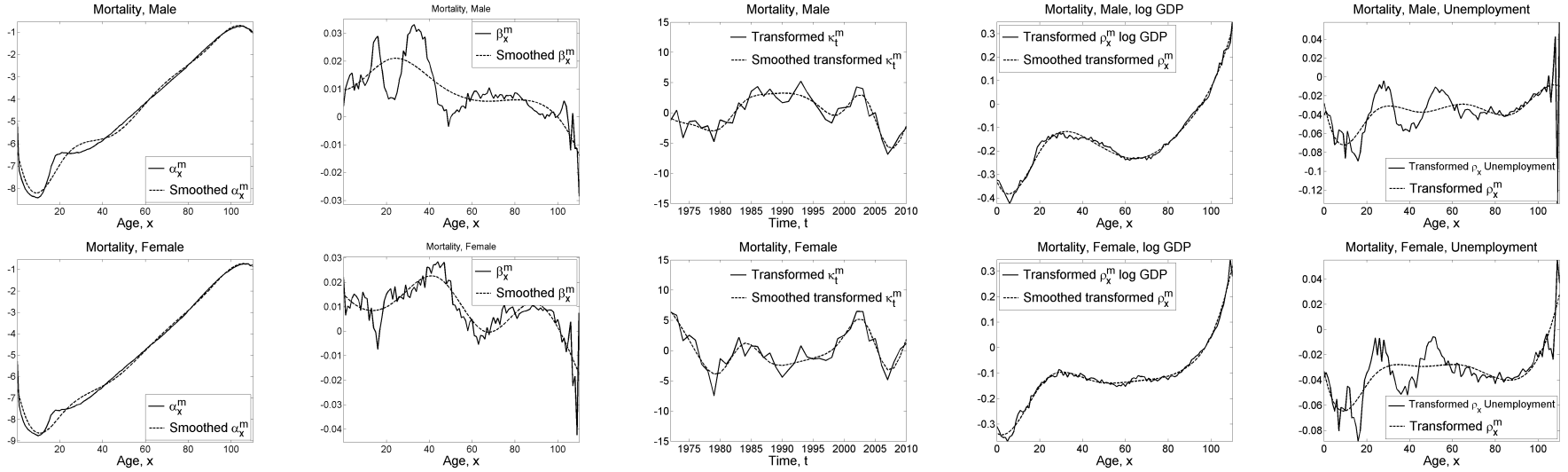

Figure 3.17 – Estimates of Lee-Carter model for mortality

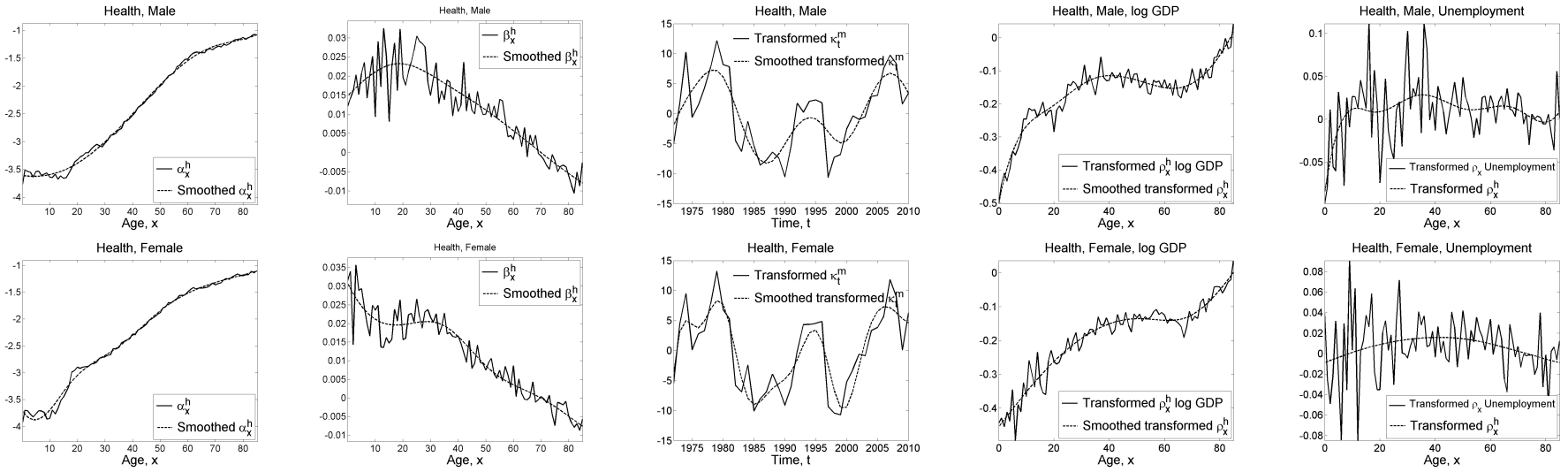

Figure 3.18 – Estimates of Lee-Carter model for health

Table 3.1 – Estimates of VAR(3) model for males, equation (4.12)

$$\boldsymbol{Y}_t \equiv \begin{bmatrix} \Delta K_t \\ \Delta Z_t \end{bmatrix} = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \Theta_2 \boldsymbol{Y}_{t-2} + \Theta_3 \boldsymbol{Y}_{t-3} + \nu_t$$

$$\hat{\boldsymbol{C}} = \begin{pmatrix} -2.78(1.05) \\ -6.79(2.44) \\ 0.02(0.05) \\ 0.29(0.35) \end{pmatrix}$$

$$\begin{bmatrix} \Delta\kappa^m_{t-1} & \Delta\kappa^h_{t-1} & \Delta\log(GDP)_{t-1} & \Delta UnEmp_{t-1} \end{bmatrix}$$

$$\hat{\Theta}_1 = \begin{pmatrix} -0.16(0.15) & 0.06(0.08) & 13.75(8.53) & 1.69(1.41) \\ -0.05(0.35) & -0.02(0.18) & -32.66(19.77) & -4.48(3.28) \\ 0.01(0.01) & -0.01(0.00) & 0.43(0.41) & 0.04(0.07) \\ -0.02(0.05) & 0.04(0.03) & -2.87(2.85) & 0.04(0.47) \end{pmatrix}$$

$$\begin{bmatrix} \Delta\kappa^m_{t-2} & \Delta\kappa^h_{t-2} & \Delta\log(GDP)_{t-2} & \Delta UnEmp_{t-2} \end{bmatrix}$$

$$\hat{\Theta}_2 = \begin{pmatrix} 0.04(0.15) & -0.09(0.07) & 2.47(6.58) & 0.75(1.15) \\ -0.72(0.34) & 0.08(0.16) & 35.00(15.24) & 6.20(2.68) \\ -0.01(0.01) & 0.00(0.00) & -0.13(0.32) & 0.02(0.06) \\ 0.00(0.05) & 0.01(0.02) & 1.97(2.20) & -0.16(0.39) \end{pmatrix}$$

$$\begin{bmatrix} \Delta\kappa^m_{t-3} & \Delta\kappa^h_{t-3} & \Delta\log(GDP)_{t-3} & \Delta UnEmp_{t-3} \end{bmatrix}$$

$$\hat{\Theta}_3 = \begin{pmatrix} 0.48(0.18) & -0.11(0.07) & 3.73(7.28) & 2.02(0.88) \\ -0.33(0.42) & -0.12(0.16) & 41.45(16.87) & 3.06(2.04) \\ 0.01(0.01) & 0.00(0.00) & 0.36(0.35) & 0.10(0.04) \\ -0.09(0.06) & 0.01(0.02) & -3.45(2.44) & -0.77(0.29) \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 1.93 & -0.23 & 0.03 & -0.23 \\ -0.23 & 10.38 & 0.03 & -0.46 \\ 0.03 & 0.03 & 0.00 & -0.03 \\ -0.23 & -0.46 & -0.03 & 0.22 \end{pmatrix}$$

Note: The table presents estimates of VAR(3) model for males with κ^m, κ^h, logarithm of GDP, and unemployment rate.
$\Delta K_t = (\Delta\kappa^m_t, \Delta\kappa^h_t)'$, $\Delta Z_t = (\Delta\log(GDP)_t, \Delta UnEmp_t)'$, where $UnEmp$ denotes unemployment rate.
Standard errors are provided in parentheses.
$\hat{\Sigma}$ is the estimated variance covariance matrix of ν.

Table 3.2 – Estimates of VAR(3) model for females, equation (4.12)

$$\boldsymbol{Y}_t \equiv \begin{bmatrix}\Delta K_t \\ \Delta Z_t\end{bmatrix} = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \Theta_2 \boldsymbol{Y}_{t-2} + \Theta_3 \boldsymbol{Y}_{t-3} + \nu_t$$

$$\hat{\boldsymbol{C}} = \begin{pmatrix} -0.83(1.13) \\ -6.41(2.59) \\ 0.02(0.04) \\ 0.41(0.31) \end{pmatrix}$$

$$\begin{bmatrix}\Delta\kappa^m_{t-1} & \Delta\kappa^h_{t-1} & \Delta\log(GDP)_{t-1} & \Delta UnEmp_{t-1}\end{bmatrix}$$

$$\hat{\Theta}_1 = \begin{pmatrix} -0.27(0.15) & 0.08(0.08) & 5.55(9.59) & 0.64(1.56) \\ -0.21(0.35) & 0.08(0.18) & -16.39(21.92) & -1.14(3.57) \\ 0.01(0.01) & -0.005(0.003) & 0.51(0.37) & 0.04(0.06) \\ -0.01(0.04) & 0.03(0.02) & -3.49(2.66) & 0.002(0.43) \end{pmatrix}$$

$$\begin{bmatrix}\Delta\kappa^m_{t-2} & \Delta\kappa^h_{t-2} & \Delta\log(GDP)_{t-2} & \Delta UnEmp_{t-2}\end{bmatrix}$$

$$\hat{\Theta}_2 = \begin{pmatrix} 0.16(0.14) & -0.08(0.08) & -8.12(8.35) & -0.96(1.37) \\ -0.22(0.31) & 0.04(0.18) & 27.40(19.09) & 5.39(3.13) \\ -0.01(0.01) & -0.005(0.003) & -0.29(0.33) & -0.01(0.05) \\ 0.03(0.04) & 0.02(0.02) & 2.50(2.32) & -0.04(0.38) \end{pmatrix}$$

$$\begin{bmatrix}\Delta\kappa^m_{t-3} & \Delta\kappa^h_{t-3} & \Delta\log(GDP)_{t-3} & \Delta UnEmp_{t-3}\end{bmatrix}$$

$$\hat{\Theta}_3 = \begin{pmatrix} 0.36(0.15) & -0.14(0.08) & 0.62(8.40) & 1.02(1.11) \\ 0.14(0.35) & -0.07(0.17) & 44.71(19.18) & 4.55(2.54) \\ 0.01(0.01) & -0.001(0.003) & 0.49(0.33) & 0.13(0.04) \\ -0.09(0.04) & -0.01(0.021) & -4.20(2.33) & -0.87(0.31) \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 2.87 & -0.60 & 0.02 & -0.22 \\ -0.60 & 15.20 & 0.04 & -0.60 \\ 0.02 & 0.04 & 0.00 & -0.03 \\ -0.22 & -0.60 & -0.03 & 0.22 \end{pmatrix}$$

Note: The table presents estimates of VAR(3) model for females with κ^m, κ^h, logarithm of GDP, and unemployment rate.
$\Delta K_t = (\Delta\kappa^m_t, \Delta\kappa^h_t)'$, $\Delta Z_t = (\Delta\log(GDP)_t, \Delta UnEmp_t)'$, where $UnEmp$ denotes unemployment rate.
Standard errors are provided in parentheses.
$\hat{\Sigma}$ is the estimated variance covariance matrix of ν.

CHAPTER 4

LINKING RETIREMENT AGE TO LIFE EXPECTANCY EFFECTS ON HEALTHY LIFE EXPECTANCY BEFORE AND AFTER RETIREMENT

This Chapter is based on De Waegenaere, Melenberg, and Yang (2014)

This chapter considers the effects of a policy that links retirement age to life expectancy. We focus on the effects on healthy life expectancy before and after retirement, and on the likelihood of being in good health at retirement age. To investigate these effects, we use a stochastic projection model that allows to jointly model and forecast health and mortality, and to quantify the corresponding uncertainties. In the best-estimate projection, linking retirement age to life expectancy would lead to an increase in retirement age of about 9 months per decade. Even though younger cohorts face significantly higher retirement ages than older cohorts, the likelihood of being in good health at retirement age is higher for younger cohorts. The effects of the policy on healthy life expectancy before and after retirement, however, are somewhat more mixed. Whereas best estimate projections suggest that healthy life expectancy before and after retirement would increase or remain constant over time, there is considerable uncertainty regarding the actual development. The bounds of the forecast intervals correspond to significant increases in time spent in poor health before retirement age, and decreases in time spent in good health after retirement age.

4.1 Introduction

The well-documented ongoing increases in life expectancy and the accompanying ageing of the population raise significant challenges for pension providers such as governments, company or private sector pension funds, and insurance companies offering life annuities (see, e.g., Bloom, Canning, Mansfield, and Moore, 2007; Hári, De Waegenaere, Melenberg, and Nijman, 2008; Pitacco, Denuit, Haberman, and Olivieri, 2009b). Moreover, a social concern of possible increasing imbalance between the growing ageing population who consume goods and services and the amount of workers who have to produce has been raised, see (see, e.g., Bloom, Canning, and Fink, 2008, 2009; Maestas and Zissimopoulos, 2010). In response to these concerns, several countries consider changes in their pension systems. A policy change that is often considered is an increase in retirement age. Several countries have already implemented (gradual) increases in retirement age, while others plan to do so in the near future. In the United States, for example, the 1983 Social Security Amendments included a provision for a gradual increase in the full retirement age for cohorts born after 1937. Specifically, retirement age has gradually increased from 65 years in 2002 to 66 years in 2008. It will stay at 66 years until 2020, and will then further increase by 2 months every year until it reaches 67 in 2026.[1] For an overview of retirement age policies in different countries, we refer to OECD (2013) and Schwan and Sail (2013).[2]

Determining a trajectory for the development of retirement age for some period in advance, however, has some potentially important drawbacks. A problem associated with this approach is that there is a considerable degree of uncertainty regarding the development of life expectancy. This has two important implications. First, there is no guarantee that the (predetermined) increase in retirement age will keep pace will the actual development of life expectancy, and so the policy might in the long run again lead to a "mismatch" between retirement age and life expectancy. Second, even though the expected level of pension payments decreases due to the policy, the degree of uncertainty in these pension payments can remain high. This uncertainty is referred to as longevity risk, and poses significant risk management challenges for governments and pension providers (see, e.g., Hári, De Waegenaere, Melenberg, and Nijman, 2008; Pitacco, Denuit, Haberman, and Olivieri, 2009b).

Because of the drawbacks associated with (long-term) deterministic retirement age policies, several countries (including Italy, Denmark, Greece, the Netherlands, Slovakia, and Cyprus) have implemented policies in which the development of retirement age is directly linked to the development of life expectancy (see, e.g., Schwan and Sail, 2013). There is some literature that investigates the benefits of such policies. Schwan and Sail (2013) investigate the budgetary impact of policies that link pension

[1] See http://www.ssa.gov/pubs/ageincrease.htm.
[2] See http://www.oecd.org/els/public-pensions/pensionsataglance.htm.

age and pension benefits to life expectancy, and find that public expenditures in the EU could almost be halved when retirement age is fully linked to life expectancy. Prettner and Canning (2013) consider a life-cycle optimization perspective to optimal retirement planning. They identify sufficient conditions under which optimal retirement age increases when life expectancy increases, and show that these conditions are satisfied for the U.S. An aspect that is ignored in these studies, however, is whether trends in health of the elderly support the raise in retirement age. Several potential concerns have been raised. First, when healthy life expectancy does not grow at the same pace as life expectancy, the policy might significantly impact the number of years that retirees can enjoy retirement in good health (see, e.g., Cutler, Meara, and Richards-Shubik, 2011). Second, concerns have been raised regarding whether people will be healthy enough to work longer (see, e.g., Munnell, Meme, Jivan, and Cahill, 2004; Munnell and Libby, 2007; Unger and Schulze, 2013). If the fraction of individuals that is not sufficiently healthy to work until retirement would increase significantly, the benefits of the policy in terms of reduced pension payments can, at least to some extent, be outweighed by increases in unemployment or other social security benefits (see, e.g., Munnell, Meme, Jivan, and Cahill, 2004). Several studies have identified such spillover effects from retirement benefits to disability benefits due to the increase in retirement age in the U.S.(see, e.g., Duggan, Singleton, and Song, 2007; Li and Maestas, 2008; Coe and Haverstick, 2010). Third, it is unclear how the policy would affect wealth transfers between males and females. Because males and females have the same retirement age even though female life expectancy is currently higher than male life expectancy, there can be significant transfers of pension between males and females. However, recent evidence suggests that female life expectancy increases at a slower rate than male life expectancy (see, e.g., Thorslund, Wastesson, Agahi, Lagergren, and Parker, 2013; Yang, De Waegenaere, and Melenberg, 2013a, and references therein). Hence, linking retirement age to gender-neutral life expectancy could non-trivially affect the gender gap, and, hence, the degree of wealth transfers.

Our focus in this paper is on the effects of a policy in which, roughly speaking, an increase in life expectancy of one month is accompanied by an increase in retirement age of one month. We investigate the effects of such a policy on life expectancy and healthy life expectancy before and after retirement, as well as the likelihood of being in good health at retirement age. We do these analyses for men and women separately. To generate projections for future mortality and health, we use a stochastic projection model based on the model developed in Yang, De Waegenaere, and Melenberg (2013a). In that paper, health and mortality are modeled using a generalization of the approach proposed by Lee and Carter (1992), and projected jointly using a Vector AutoRegressive approach. An important advantage of this method is that it not only generates point forecasts but also the corresponding forecast uncertainty. Moreover, it allows to take into account dependence between the developments of mortality and health. Be-

cause for our purpose also dependence between the developments of gender-neutral mortality rates (which form the basis of the retirement age decision) and the gender-specific male and female mortality rates (which determine how the policy affects males and females) is important, we slightly extend the model in Yang, De Waegenaere and Melenberg (2013) by jointly modeling mortality and health rates of both genders. The model is estimated on mortality and self-assessed health data for the U.S. male and female population in the period 1972-2010.

Our results suggest that linking retirement age to life expectancy would lead to significant increases in retirement age; best-estimate projections suggest an increase of on average 9 months per decade. Even though retirement age increases quite sharply, the probability of being in good health at retirement age increases significantly. These results suggest that improvements in health indeed support the increase in retirement age. The results regarding the number of years spent in good health before and after retirement are a bit mixed. In the best-estimate scenario, linking retirement age to gender-neutral life expectancy would lead to relatively constant levels of both life expectancy and healthy life expectancy after retirement, for both males and females. However, there is a considerable degree of uncertainty. Whereas the best-estimate projections suggest that all future cohorts would enjoy an almost equal number of healthy years in retirement, the bounds of the forecast intervals present a more pessimistic view. At these bounds, the time spent in good health during retirement would decrease by a little bit less than five months per decade, for both males and females. Moreover, the bounds of the forecast intervals that we generate correspond to an increase in the time spent in poor health between age 60 and retirement age by 2 months per decade for males, and by 1.4 months per decade for females. Overall, these results suggest that focusing on best-estimate projections, as is often done by policy makers, might provide a too optimistic view regarding the feasibility of retirement age policies that link retirement age to life expectancy.

The rest of this paper is organized as follows. Section 4.2 defines the retirement age policy that is considered in this paper, and presents the model that is used to generate forecasts for mortality and health. Section 4.3 presents the empirical results on the development of retirement age, life expectancy and healthy life expectancy before and after retirement, and the probability of being in good health at retirement age. We conclude in Section 5.5.

4.2 Model and methods

Our goal in this paper is to investigate the effect of a policy that links retirement age to life expectancy on healthy life expectancy before and after retirement, and on the likelihood of being in good health at retirement age. In this section, we present the model and methods used throughout the paper. In Section 4.2.1, we define the retirement age

policy that will be investigated in this paper. In Section 4.2.2, we briefly discuss the method used to forecast mortality and health. More details regarding this method can be found in Yang, De Waegenaere, and Melenberg (2013a).

4.2.1 Retirement age policy

We consider a policy in which retirement age is directly linked to the development of life expectancy. The goal of the policy is to adjust retirement age in such a way that all future cohorts face approximately the same expected number of years lived after retirement.

There are some practical complications associated with linking retirement age to remaining life expectancy. First, remaining life expectancy for an individual with a given age x depends on death rates at ages higher than x. Unless one is considering a cohort for which all individuals have died (e.g., the cohort born in 1880), death rates for future years are needed in order to determine the individual's remaining life expectancy. There is, however, a significant degree of uncertainty regarding the development of mortality rates, and projections for future mortality rates are sensitive to model risk. To avoid this model risk, policies that link retirement age to life expectancy are typically based on changes in *period* life expectancy rather than based on changes in *cohort* life expectancy. Period life expectancy in a given year is determined under the assumption that death rates for all future years are identical to the death rates in the last observed year. Using period life expectancy to adjust retirement age therefore has the important advantage that the determination of life expectancy is not sensitive to model risk, since it is based on observed mortality rates only. A potential drawback of linking retirement age to changes in period life expectancy is that the policy might not lead to constant life expectancy at retirement age, if changes over time in period life expectancy deviate significantly from changes over time in cohort life expectancy. We discuss this issue in detail in Section 4.3.2. A second practical complication associated with linking retirement age to life expectancy is that life expectancy depends significantly on gender. In many countries, however, it is prohibited to differentiate between men and women in terms of retirement age. Therefore, we consider a policy in which the adjustment of retirement age is based on *gender neutral* period life expectancy. This implies that life expectancy is determined based on population mortality rates, rather than gender-specific mortality rates.

To formally define the retirement age policy that will be investigated in this paper, we introduce the following notation:

- $LE^{period}(x,t)$: gender-neutral period life expectancy of an x-year old in year t; the expression for $LE^{period}(x,t)$ is presented in Appendix 4.A (in (4.26));
- t_0: the year in which the policy is first implemented;

- $RetAge(t)$: the retirement age in year t, where we consider only retirement ages that are multiples of years and whole months.

If the goal is to adjust retirement age in such a way that (gender-neutral period) life expectancy after retirement is constant, then retirement age in year t should be determined such that:

$$LE^{period}(RetAge(t), t) = LE^{period}(RetAge(t_0), t_0). \tag{4.1}$$

This, however, implies that determining retirement age requires solving an equation (i.e., (4.1)). For that reason, most countries choose for a somewhat simpler policy that yields an explicit rather than implicit expression for retirement age. In this paper, we focus on a policy similar to the policy implemented in the Netherlands in 2013. The policy determines as benchmark age the retirement age at the start of the implementation of the policy, i.e., $RetAge(t_0)$. In any future year, one determines by how many months remaining (gender neutral period) life expectancy in year t at the benchmark age has increased as compared to year t_0. We denote this increase by $\Delta LE^{period}(t)$, i.e.,

$$\Delta LE^{period}(t) = \left\lfloor LE^{period}(RetAge(t_0), t) - LE^{period}(RetAge(t_0), t_0) \right\rfloor_{months}, \tag{4.2}$$

where for any x, $\lfloor x \rfloor_{months}$ means that x is rounded below to a multiple of $1/12$. Because decreases in retirement age over time are ruled out, retirement in year $t = t_0 + 1, \cdots, T$, retirement age is determined as:

$$RetAge(t) = \max\left\{ RetAge(t-1), RetAge(t_0) + \Delta LE^{period}(t) \right\}. \tag{4.3}$$

For any given scenario for the development over time of gender neutral life expectancy, the corresponding development over time of retirement age can be determined from (4.2) and (4.3). In the next section, we discuss how scenarios for gender-neutral death rates are generated.

4.2.2 Projecting mortality and health

Our focus is on the effects of the retirement age policy defined in Section 4.2.1 on (healthy) life expectancy before and after retirement, and on the likelihood of being in good health at retirement age. To determine life expectancy and healthy life expectancy before and after retirement age, we need projections of death rates and health status as a function of age, for both males and females. Moreover, for the determination of retirement age, we need projections of gender neutral death rates. In Sections 4.2.2 and 4.2.2, we first discuss how we measure mortality and health. The model that we use to forecast the future developments of health and mortality is presented in Section

4.2.2. In Section 4.2.2, we discuss the data. The parameter estimates of the model are presented in Appendix 4.B.

Mortality measurement

In this section, we discuss the measures that we use for mortality, as a function of age and gender. We use the following notation:

- $g \in \{m, f, n\}$: an indicator for the gender specification, where $g = m$ refers to males, $g = f$ refers to females, and $g = n$ refers to the gender neutral case;
- ω: the maximum attainable age; we assume that any individual who reaches age ω dies within the next month;
- $\mathcal{A} = \{0, 1, \ldots, \omega - 1\}$: the set of integer ages, except for the maximum attainable age ω.

For mortality, we model the development of the *central death rate*, which is defined as the ratio of the number of deaths at age x in year t for gender g, denoted $D_{x,t}^{(g)}$, to the exposure to death at age x in year t for gender g, denoted $E_{x,t}^{(g)}$, i.e.,

$$m_{x,t}^{(g)} = \frac{D_{x,t}^{(g)}}{E_{x,t}^{(g)}}, \text{ for } x \in \mathcal{A} \text{ and } g \in \{m, f\}. \tag{4.4}$$

(see, e.g., Chapter 3 in Pitacco, Denuit, Haberman, and Olivieri (2009b)). For any given scenario for the development of the central death rates $m_{x,t}^{(g)}$ for males and females, the corresponding one-year death probabilities can be determined from (see, e.g., Molla, Wagener, and Madans (2001))

$$q_{x,t}^{(g)} = \frac{m_{x,t}^{(g)}}{1 + m_{x,t}^{(g)}}, \tag{4.5}$$

for all ages $x \in \mathcal{A}$, and for both genders $g \in \{m, f\}$.

Because retirement age can be expressed as multiples of months, we transform the one-year death probabilities into monthly death probabilities. To transform one-year probabilities to monthly death probabilities, we assume that for any given $x \in \mathcal{A}$, the probability of dying in the age interval $[x + s, x + s + 1)$ is independent of s for $s \in \{0, \frac{1}{12}, \ldots, \frac{11}{12}\}$. Hence, the gender-specific probability of dying within the next month for an individual aged $x + s$ at time $t + s$ is given by:

$${}_{\frac{1}{12}}q_{x+s,t+s}^{(g)} = 1 - \left(1 - q_{x,t}^{(g)}\right)^{\frac{1}{12}}, \tag{4.6}$$

for all ages $x \in \mathcal{A}$, all $s \in \{0, \frac{1}{12}, \ldots, \frac{11}{12}\}$, and both genders $g \in \{m, f\}$. Moreover, because retirement age is based on *gender-neutral* life expectancy, we need gender-neutral monthly death probabilities. Gender-neutral monthly death probabilities are determined as the average of the male and female one-year death probabilities, i.e.,

$$ {}_{\frac{1}{12}}q^{(n)}_{x+s,t+s} = \left({}_{\frac{1}{12}}q^{(m)}_{x+s,t+s} + {}_{\frac{1}{12}}q^{(f)}_{x+s,t+s} \right) / 2, \tag{4.7} $$

for all ages $x \in \mathcal{A}$, and all $s \in \{0, \frac{1}{12}, \ldots, \frac{11}{12}\}$.[3]

Health measurement

To measure health, we use self-assessed health information. The Integrated Health Interview Series (IHIS) integrates survey data collected by the National Health Interview Survey (NHIS), and provides consecutive annual cross-sectional self-assessed health status at the national level , distinguishing "excellent", "good", "fair", and "poor" before 1982, and "excellent", "very good", "good", "fair", and "poor"after 1982. The NHIS considers individuals in the age groups $x \in \{0, 1, \ldots, 85+\}$, where the age group 85+ consists of all individuals aged 85 or higher.

As in Yang, De Waegenaere, and Melenberg (2013a), survey respondents who report their health status as "poor" and "fair" are clustered in the "unhealthy" group, and survey respondents who report the other three better health categories are clustered in the "healthy" group. Hence, we define the *health status index* of x-year olds in year t with gender g as the fraction who is "unhealthy" (see, e.g., Molla, Wagener, and Madans (2001)):

$$ \hat{\pi}^{(g)}_{x,t} = \frac{1}{\sum_{j=1}^{N^{(g)}_{x,t}} w^{(g)}_{j,x,t}} \sum_{j=1}^{N^{(g)}_{x,t}} w^{(g)}_{j,x,t} H^{(g)}_{j,x,t}, \text{ for } x \in \{0, 1 \ldots, 84, 85+\} \text{ and } g \in \{m, f\}, \tag{4.8} $$

where $N^{(g)}_{x,t}$ represents the total number of survey respondents in age group x in year t, $H^{(g)}_{j,x,t}$ is a zero-one indicator, with $H^{(g)}_{j,x,t} = 1$ if the self-assessed health of the j^{th} respondent with age x was either fair or poor, and $H^{(g)}_{j,x,t} = 0$ otherwise, and $w^{(g)}_{j,x,t}$ is the person weight, representing the inverse probability of the j^{th} respondent in the population selected into the sample. Section 3.1 in Yang, De Waegenaere, and Melenberg (2013b) provides a more detailed description of the self-assessed health data, including

[3]This way of determining gender neutral probabilities corresponds to how these probabilities are often determined in practice. From a theoretical perspective, however, one would instead determine the gender neutral probabilities by determining $m_{x,t}$ based on death numbers, $D_{x,t}$, and exposures to death, $E_{x,t}$, of males and females together. Given the focus of our analysis, however, we use the approach that is most often used in practice.

response rates and person weight.

To compute the healthy life expectancy, we need values of the health status indicator $\pi_{x,t}^{(g)}$ for each individual age $x \in \{85, \ldots, \omega\}$ in the age group 85+. Following Yang, De Waegenaere, and Melenberg (2013a), we set[4]

$$\pi_{x,t}^{(g)} = \pi_{85+,t}^{(g)}, \text{ for } x = 85, 86, \ldots, \omega.$$

As was the case for mortality rates, we need to transform the yearly health status index to a monthly health status index in order to determine healthy life expectancy at non-integer ages. To do so, we assume that the health status index in the age interval $[x, x+1)$ increases linearly from $\pi_{x,t}^{(g)}$ to $\pi_{x+1,t+1}^{(g)}$, i.e., we assume that:

$$\pi_{x+s,t+s}^{(g)} = \pi_{x,t}^{(g)} + s\left(\pi_{x+1,t+1}^{(g)} - \pi_{x,t}^{(g)}\right), \tag{4.9}$$

for all ages $x \in A$, all $s \in \{0, \frac{1}{12}, \ldots, \frac{11}{12}\}$, and both genders $g \in \{m, f\}$.

Forecast model

The forecast model that we use, models the development over time of the yearly central death rates $m_{x,t}^{(g)}$, and yearly health status indices $\pi_{x,t}^{(g)}$, for both genders. For any given scenario for $m_{x,t}^{(g)}$ and $\pi_{x,t}^{(g)}$, monthly death probabilities and health status indices can be determined from (4.5)-(4.9).

In order to take into account the likely dependence between mortality and health, we use the forecast method proposed in Yang, De Waegenaere, and Melenberg (2013a), in which mortality rates and health prevalence rates are modeled and projected jointly. Because for our purpose also dependence between the developments of gender-neutral mortality rates and gender-specific male and female mortality rates is important, we slightly extend the model in Yang, De Waegenaere and Melenberg (2013) by jointly modeling mortality and health rates of both genders. In this section, we briefly summarize the approach. For a more detailed discussion, we refer to Yang, De Waegenaere, and Melenberg (2013a).

In the first step, the developments of mortality and health are modeled separately, using a generalization of the approach proposed by Lee and Carter (1992) for mortality. Whereas Lee and Carter (1992) models mortality rates as a function of latent variables only, Yang, De Waegenaere, and Melenberg (2013a) model both mortality and health, and include in addition to the latent variables also GDP and unemployment rate as explanatory variables. They find that including these two macroeconomic variables leads

[4]When quantifying the corresponding confidence intervals, we explicitly take into account that there is more uncertainty in the value of $\pi_{x,t}$ for higher ages than for lower ages. This higher uncertainty is reflected by a higher variance in the measurement errors $\epsilon_{x,g,t}^{h}$, in equation (4.11). For further details, we refer to Yang, De Waegenaere, and Melenberg (2013a)

to significant improvements in both the model fit (as measured by mean square error) and the forecast accuracy of the model (as measured by the mean squared forecasting error). For each gender specification $g \in \{m, f\}$ separately, the dynamics of $m_{x,t}^{(g)}$ and $\pi_{x,t}^{(g)}$ are modeled as follows:

$$\log\left(m_{x,t}^{(g)}\right) = \alpha_{x,g}^{m} + \beta_{x,g}^{m} \cdot \kappa_{t,g}^{m} + \left(\rho_{x,g}^{m}\right)' \cdot Z_t + \epsilon_{x,g,t}^{m}, \tag{4.10}$$

$$\log\left(\pi_{x,t}^{(g)}\right) = \alpha_{x,g}^{h} + \beta_{x,g}^{h} \cdot \kappa_{t,g}^{h} + (\rho_{x,g}^{h})' \cdot Z_t + \epsilon_{x,g,t}^{h}. \tag{4.11}$$

For both the mortality model ($c = m$) and the health model ($c = h$), and for both genders, $\alpha_{x,g}^{c}$ describes the age-specific time-independent level of mortality (health); $\kappa_{t,g}^{c}$ is a time-dependent univariate latent variable that reflects the change in the overall level of mortality (health) over time; $\beta_{x,g}^{c}$ reflects the age-specific sensitivity of mortality (health) to changes in $\kappa_{t,g}^{c}$; Z_t is a 2-dimensional vector containing log GDP and unemployment rate in year t; $\rho_{x,g}^{c}$ is a 2-dimensional age-specific parameter vector that reflects the effect of changes in log GDP and unemployment rate on the overall level of mortality (health), and $\epsilon_{x,g,t}^{c}$ is the error term, reflecting idiosyncratic time- and gender-specific influences, with mean 0 and (possibly age-specific) variance $\sigma_{\epsilon,x}^{2}$.

In the second step, the joint dynamics of the latent time trends for mortality and health of both genders ($\kappa_{t,g}^{m}$ and $\kappa_{t,g}^{h}$, for $g \in \{m, f\}$), as well as the two observed variables (log GDP and unemployment rate), are modeled and projected using Vector AutoRegression (VAR). The approach that we use in this paper deviates from that in Yang, De Waegenaere, and Melenberg (2013a) in one aspect. Yang, De Waegenaere, and Melenberg (2013a) estimate a separate VAR model for males and females. For our purpose, however, joint modeling of male and female mortality and health (taking into account correlations between males and females) is important for two reasons. First, ignoring the correlation between male and female mortality will likely lead to biased forecasts of the gender-neutral measure of life expectancy that is used to determine retirement age (see (4.3)). Second, biases in the correlation between male and female health and/or mortality lead to biases in the correlation between retirement age (which is based on gender-neutral measure of life expectancy) and gender-specific life expectancy and healthy life expectancy before and after retirement. To avoid these biases, we extend the model in Yang, De Waegenaere, and Melenberg (2013a) to allow for joint modeling and projecting of male and female mortality and health. We choose the lag length in the VAR model according to the Akaike (AIC), Bayesian (BIC), and Hannan-Quinn (HQIC) information criteria. These information criteria provide a trade-off between goodness-of-fit and the parsimony of the model. Because we have a relatively small sample size (i.e., 39 years), and the number of variables included in the VAR model is 6, we set the maximum tested lag length equal to 3. Both AIC and HQIC suggest a lag length of 3, whereas BIC suggests a lag length of 0. Lütkepohl (2007)

compares these information criteria when selecting the lag length in a VAR model, and finds that in settings with moderate or small sample sizes, AIC is preferred over BIC. As our sample size is relatively small, we set the lag length equal to 3. This results in the following model.

$$\boldsymbol{Y}_t \equiv \begin{pmatrix} \Delta K_t \\ \Delta Z_t \end{pmatrix} = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \Theta_2 \boldsymbol{Y}_{t-2} + \Theta_3 \boldsymbol{Y}_{t-3} + \nu_t, \tag{4.12}$$

where K_t is a vector with components $(\kappa^m_{t,m}, \kappa^m_{t,f}, \kappa^h_{t,m}, \kappa^h_{t,f})$, ΔK_t and ΔZ_t are the first order differences of K_t and Z_t, i.e., $\Delta K_t \equiv K_t - K_{t-1}$, and $\Delta Z_t \equiv Z_t - Z_{t-1}$, and ν_t is a six-dimensional vector of white noise terms with covariance matrix Σ_ν.

In our forecasts, we consider two sources of risk. First, we include process risk that arises due to uncertainty in ν_t in (4.12), as well as $\epsilon^c_{x,t}$ in (4.10) and (4.11) for both mortality ($c = m$) and health ($c = \pi$). In addition to this, we consider parameter risk, which arises due to uncertainty is the estimated parameters in (4.12). A standard bootstrap procedure is used to generate scenarios for these parameters. Readers can refer to Yang, De Waegenaere, and Melenberg (2013a) for a detailed discussion on the method of predicting mortality and health prevalence and quantifying their uncertainties.

Data and parameter estimates

As in Yang, De Waegenaere, and Melenberg (2013a), we use the following data, for the period 1972 to 2010:[5]

- consecutive annual cross-sectional gender-specific death numbers, $D^{(g)}_{x,t}$, and gender-specific exposures to death, $E^{(g)}_{x,t}$, for the U.S. population, for males and females, obtained from the Human Mortality Database (HMD);[6]
- consecutive annual cross-sectional self-assessed health for the noninstitutionalized population in the U.S., for males and females, obtained from the NHIS;[7]
- Gross Domestic Product (GDP) per capita, measured in U.S. dollar constant prices (OECD base year 2005). Data is obtained from the Organisation for Economic Cooperation and Development (OECD) website.[8]
- total unemployment rate, obtained from the OECD website (Statistics Extracts, Country Statistical Profiles, 2010).[9]

[5]This is the longest period for which self-assessed health data is available.
[6]See http://www.mortality.org/
[7]See http://www.ihis.us/ihis/
[8]See http://stats.oecd.org/
[9]OECD Statistics defines the total unemployment rate as the ratio of people aged 15 years and over that are out of work and actively seeking it to the population of working age either in work or actively seeking it, all raw and seasonally adjusted.

For a more detailed description of the data, we refer to Yang, De Waegenaere, and Melenberg (2013a). Estimates of the parameters of the mortality and health models in (4.10) and (4.11), and of the parameters of the $VAR(3)$ model in (4.12) are presented in Appendix 4.B.

4.3 Effects of the retirement age policy

In this section, we analyze the effects of the retirement age policy as defined in (4.13) on (healthy) life expectancy before and after retirement, and on the probability of being in good health at retirement date. If the policy would imply that the expected number of years in poor health before retirement increases significantly, or that the fraction of individuals that is in good health at retirement age decreases significantly, this could be an indication that fewer people will be willing to, or able to, work until retirement age when retirement age is linked to life expectancy. This, in turn, could imply that the policy would have adverse effects on unemployment benefits or disability benefits. Moreover, if the policy would imply that the expected number of years lived in good health after retirement decreases significantly, this could decrease the support for the policy, as it could be perceived as unfair.

Throughout the paper, we use the following notation:

- $LE^{cohort}(x, y, g, t)$: the gender-specific expected number of years lived between ages x and y, for x-year olds in year t with gender $g \in \{m, f, n\}$, conditional on being alive at age x;
- $HLE^{cohort}(x, y, g, t)$: the gender-specific expected number of years lived in good health between ages x and y, for x-year olds in year t with gender $g \in \{m, f, n\}$, conditional on being alive at age x.

The (healthy) life expectancy measures $LE^{cohort}(x, y, g, t)$ and $HLE^{cohort}(x, y, g, t)$ are based on projected gender-specific *cohort* monthly life tables. In contrast to the period life tables used to determine retirement age, cohort life tables take into account projected future changes in mortality rates. Expressions for $LE^{cohort}(x, y, g, t)$ and $HLE^{cohort}(x, y, g, t)$ are given in (4.27) and (4.28) in Appendix 4.A.

This section is organized as follows. In Section 4.3.1, we illustrate the effect of the policy on the development of retirement age as a function of time, as well as a function of the birth cohort. We also compare the corresponding retirement age policy to the policy implemented under the 1983 Social Security Amendments. In Section 4.3.2, we show the effect on (healthy) life expectancy after retirement for males and females. In Section 4.3.3, we analyze the effect of the policy on the likelihood of being in good health at retirement age, as well as on the expected number of years lived in poor health between age 60 and retirement age. All results are based on 2000 scenarios for

the developments of $m_{x,t}^{(g)}$ and $\pi_{x,t}^{(g)}$, for males and females, generated with the model described in Section 4.2.2.

4.3.1 Retirement age

In this section we show the effect of the policy defined in Section 4.2.1 on the development of retirement age over time. We first determine the development of retirement age as a function of time, as well as a function of the birth cohort. Then, we compare the retirement ages according to the 1983 Social Security Amendments to the retirement age suggested by the policy that we consider.

Because retirement age in 2010 in the United States is 66 years, we let $t_0 = 2010$ and $RetAge(t_0) = 66$. Moreover, remaining life expectancy of a 66 year old in 2010, based on the period life table in 2010, equals 18.58 years. Hence, according to the policy defined in (4.3), retirement age in year t would be determined by:

$$RetAge(t) = \max\left\{RetAge(t-1), 66 + \left\lfloor LE^{period}(66,t) - 18.58 \right\rfloor_{months}\right\}, \tag{4.13}$$

for $t = 2011, \ldots, 2060$. To determine retirement age for different cohorts, we introduce the following notation:

- $t(x)$: the year in which an x-year old in year t_0 will retire;
- $R(x)$: the age at which an x-year old in year t_0 will retire.

Now consider an individual belonging to the cohort aged x-years old in year t_0 who is not yet retired at the beginning of year t. This individual will reach retirement age during year t if $RetAge(t) - (x + t - t_0) < 1$. Hence, the year in which (s)he retires, $t(x)$, is given by:

$$t(x) = \min\left\{t \in \{t_0, t_0+1, \cdots, T\} : RetAge(t) - (t - t_0 + x) < 1\right\}, \tag{4.14}$$

and the corresponding retirement age is given by

$$R(x) = RetAge(t(x)). \tag{4.15}$$

For each of the 2000 scenarios for the development over time of the central death rates $m_{x,t}^{(g)}$ for males and females, (using the model described in Section 4.2.2), the corresponding retirement age in a future year t in that scenario is determined from (4.13), using (4.5), (4.7), and (4.26). We determine both the best-estimate projection, as well as the 95% forecast interval, where for the latter we distinguish the case where only process risk is taken into account, and the case where both process risk and parameter risk are taken into account.

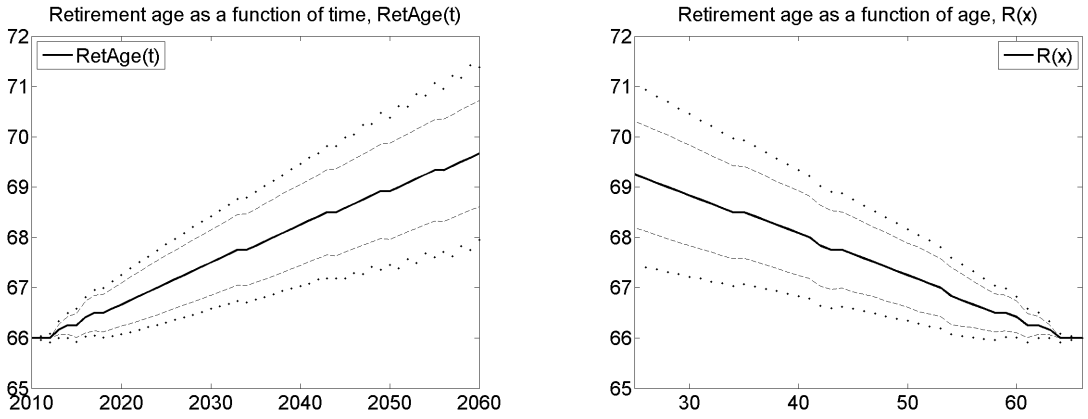

Figure 4.1 – Retirement age as a function of time ($RetAge(t)$, left panel) for $t = 2010, \ldots, 2060$, and as a function of age in 2010 for $x = 25, \ldots, 66$ ($R(x)$, right panel). In each figure, the solid line represents the best-estimate forecast; the dashed and the dotted lines represent the 95% forecast intervals, where the smaller intervals reflect process risk only while the wider intervals reflect both process risk and parameter risk.

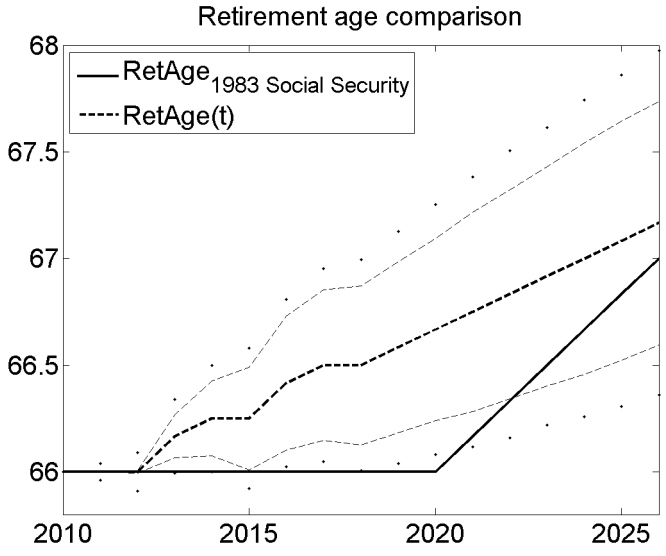

Figure 4.2 – Retirement age according to the U.S. 1983 Social Security Amendments (solid line), and according to the retirement age policy in (4.13) (dashed lines and dotted lines), for $t = 2010, \ldots, 2026$.

The left panel in Figure 4.1 displays the expected value of retirement age in year t, i.e., $RetAge(t)$ for $t = 2011, 2012, \ldots$,2060. The right panel shows the retirement age as a function of the birth cohort, i.e., $R(x)$ for cohorts aged $x = 25, \ldots, 66$, in 2010. In each figure, the solid line represents the best-estimate forecast; the dashed lines represent the 95% forecast intervals, where the smaller intervals reflect process risk only while the wider intervals reflect both process risk and parameter risk. The best-estimate is determined by setting $\epsilon^m_{x,g,t} = 0$ in (4.10) and $\epsilon^h_{x,g,t} = 0$ in (4.11), and $\nu_t = 0$ in (4.12). The figure shows an increase in the best-estimate projected retirement age from 66 in 2010 to 69 years and 8 months in 2060. However, there is a significant degree of uncertainty regarding the development of retirement age. When both process risk and parameter risk are taken into account, the upper bound of the 95% forecast interval in 2060 is 71 years and 4 months; the lower bound is 67 years and 11 months.

We conclude this section by comparing the retirement age according to the 1983 Social Security Amendments to the retirement age according to the policy that we investigate. The 1983 Social Security Amendments implies that retirement age gradually increases from 65 for individuals born in 1937 or earlier to 67 for individuals born after 1960 and later. Figure 4.2 shows both the retirement age according to the 1983 Social Security Amendments and the best-estimate retirement age according to the policy that we investigate, as a function of the year in which the individual retires. Because the 1983 Social Security Amendments defines retirement age until $t = 2026$, we consider only these years in our comparison.

The figure shows that linking retirement age to life expectancy according to (4.13) yields a significantly higher retirement age in $t = 2013, \ldots, 2021$ than the one that is implemented in the 1983 Social Security Amendments. Because the policy that we consider aims at adjusting retirement age in such a way that remaining life expectancy after retirement stays constant, this suggests that under the current U.S. policy, remaining life expectancy after retirement age will continue to increase over time. In later years, the 95% forecast interval contains both scenarios in which the retirement age according to (4.13) is higher and scenarios in which it is lower than under the current policy.

4.3.2 Life expectancy and healthy life expectancy after retirement

In this section we show the effect of the policy defined in (4.3) on life expectancy and healthy life expectancy after retirement, for different cohorts. Although the general idea behind linking retirement age to life expectancy as in (4.3) is to keep the number of years spent in retirement stable over time, there are several reasons why this is not guaranteed. First, keeping retirement age constant over time would require retirement age to be determined from (4.1), which is slightly different from how retirement age is determined under the policy, i.e., via (4.3). Second, as discussed in Section 4.2.1, the

policy is based on period life tables, which implies that changes in mortality rates after the date at which life expectancy is determined are not taken into account. Hence, when mortality rates decline over time, period life expectancy is an underestimate of the cohort's "true" life expectancy. This implies that period life expectancy might not accurately capture the actual change in life expectancy of different cohorts. Third, to avoid gender "discrimination", the measure of life expectancy used to determine retirement age is based on gender-neutral mortality rates. It is well-documented, however, that males and females not only have significantly different life expectancies, but that the trends in these life expectancies over time are also different (see, e.g., Thorslund, Wastesson, Agahi, Lagergren, and Parker (2013); Yang, De Waegenaere, and Melenberg (2013a), and references therein). Because retirement age is based on gender neutral life expectancy, there is no guarantee that gender-specific life expectancy remains constant even if it were stable at population level. In order to investigate the potential differential effect of the policy on men and women, we consider gender-neutral as well as gender-specific (healthy) life expectancies.

Because retirement age in year t equals $RetAge(t)$, and because the maximum attainable age in our model is 110, the remaining life expectancy of an individual with gender g who retires in year t equals

$$LE_{retiree}(t,g) = LE^{cohort}(RetAge(t), 110, g, t), \tag{4.16}$$

and the remaining healthy life expectancy of an individual with gender g who retires in year t equals

$$HLE_{retiree}(t,g) = HLE^{cohort}(RetAge(t), 110, g, t), \tag{4.17}$$

where LE^{cohort} and HLE^{cohort} are as defined in (4.27) and (4.28). For each of the 2000 scenarios for the development over time of the central death rates ($m^{(g)}_{x,t}$) and health status index ($\pi^{(g)}_{x,t}$) for males and females, the corresponding retirement age in a future year t in that scenario is determined from (4.13), using (4.5), (4.7), and (4.26), and (healthy) life expectancy at retirement age can be determined from (4.16) and (4.17), using (4.9), (4.25), (4.27), and (4.28).

The left panel in Figure 4.3 shows the best-estimate projections of $LE_{retiree}(t,g)$, for all population (solid lines, $g = n$), for males (dashed lines, $g = m$) and for females (dotted lines, $g = f$).

The best-estimate projections provide an optimistic view for policymakers. First, remaining life expectancy at retirement is relatively stable over time for both males and females, with a very slight decreasing trend for females and a somewhat stronger but still small increasing trend for males. This occurs because life expectancy for females increases at a slower rate than gender-neutral life expectancy, and, hence, the opposite

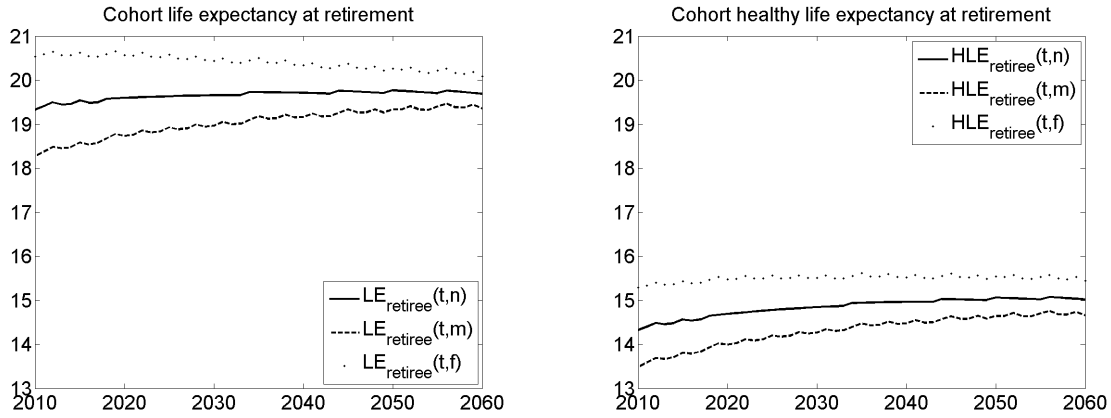

Figure 4.3 – Best estimate projections of life expectancy after retirement (left panel), and healthy life expectancy after retirement (right panel), for individuals who retire in year $t = 2010, \ldots, 2060$. The solid lines represent gender-neutral (healthy) life expectancy, the dotted lines represent female (healthy) life expectancy, and the dashed lines represent male (healthy) life expectancy.

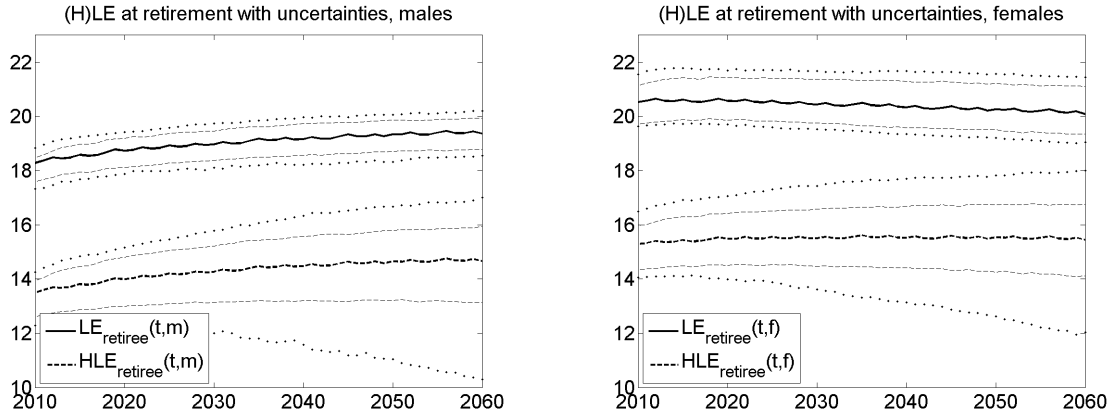

Figure 4.4 – Best estimate projections and 95% forecast interval of life expectancy and healthy life expectancy after retirement for males (left panel) and for females (right panel), for individuals who retire in year $t = 2010, \ldots, 2060$. In each figure, the solid line represents the best-estimate forecast; the dashed lines and dotted lines represent the 95% forecast intervals, where the smaller intervals reflect process risk only while the wider intervals reflect both process risk and parameter risk.

occurs for males. Because retirement age is based on the trend in gender-neutral life expectancy, this implies that the expected number of years spent in retirement decreases slightly over time for females, and increases slightly over time for males. Even though these effects are small, they do suggest that in the long run the gender disparity at retirement will decrease. While the best-estimate of the difference in remaining lifetime after retirement between males and females is almost 2.27 years in 2010, it is projected to decrease to 0.73 years in 2060. Second, best-estimate projections for healthy life expectancy at retirement for both males and females are relatively stable over time. This suggests that the increase in healthy life expectancy keeps pace with the increase in life expectancy, so that the policy would not adversely affect the number of years spent in good health after retirement. However, these conclusions are based exclusively on best-estimates, i.e., they do not account for the relatively large uncertainty in the developments of mortality and health. While linking retirement age to life expectancy yields a significant reduction in the degree of uncertainty in remaining life expectancy after retirement, the uncertainty is not fully "hedged", for two reasons. First, as discussed above, the measure of life expectancy used by the policy deviates from the cohort's true life expectancy. Second, after retirement age is set, there can still be developments in life expectancy. To quantify this remaining uncertainty, we generated forecast intervals for (healthy) life expectancy, taking into account both process risk and parameter risk, as discussed in Section 4.2. Figure 4.4 displays the developments of remaining lifetime and remaining lifetime in good health at retirement for males and females, including the 95% forecast intervals. In each graph, the narrower forecast intervals reflect only process risk, whereas the wider forecast intervals reflect both process and parameter risks.

Figure 4.4 shows that the degree of uncertainty in remaining life expectancy after retirement is relatively small. The degree of uncertainty regarding the development of healthy life expectancy after retirement is much larger. For males, the remaining healthy life expectancy at retirement for the cohort that retires in 2060 has a lower bound of 10.30 years and an upper bound of 17.00 years, whereas the best estimate of healthy life expectancy in 2010 is 13.48 years. This implies that there is a significant probability that the number of years in retirement spent in good health will decrease substantially if retirement age is linked to life expectancy. For females, there is a similar degree of uncertainty.

4.3.3 Healthy enough to work until retirement?

As mentioned in the introduction, a potential concern associated with policies that link retirement age to life expectancy is that they might have a significant impact on the fraction of individuals that remain in relatively good health until retirement. If that fraction decreases significantly, this could be an indication that the policy could ad-

versely affect unemployment or disability payments (see, e.g., Munnell, Meme, Jivan, and Cahill, 2004; Duggan, Singleton, and Song, 2007; Li and Maestas, 2008; Coe and Haverstick, 2010).

In this section, we focus on the effect of the retirement age policy on the number of years lived in poor health between age 60 and retirement age (Section 4.3.3), and on the likelihood of being in good health at retirement age (Section 4.3.3). Because empirical literature suggests that individuals who report poor self-assessed health are more likely to retire early (see, e.g., McGarry, 2004; Jones, Rice, and Roberts, 2010), a higher value of the expected number of years lived in poor health or a lower likelihood of being in good health at retirement age can be indicators that the fraction of individuals that may not be able to, or willing to, work until extended retirement age could increase due to the policy.

Time spent in poor health before retirement age

In this section we show the effect of the retirement age policy defined in (4.3) on the expected number of years spent in poor health between age 60 and retirement age, as this could be an indicator of the extent to which individuals will be able (or willing) to work until the extended retirment age.

Because retirement age of an individual aged 60 in year t equals $R(2010 - t + 60)$, the expected years lived between age 60 and retirement for that individual equals

$$LE_{active}(t, g) = LE^{cohort}(60, R(60 - (t - 2010)), g, t), \tag{4.18}$$

and the expected number of years lived in good health between age 60 and retirement equals

$$HLE_{active}(t, g) = HLE^{cohort}(60, R(60 - (t - 2010)), g, t). \tag{4.19}$$

For each of the 2000 scenarios for the development over time of the central death rates ($m^{(g)}_{x,t}$) and health status index ($\pi^{(g)}_{x,t}$) for males and females, the corresponding retirement age in a future year t in that scenario is determined from (4.13), using (4.5), (4.7), and (4.26), and (healthy) life expectancy at retirement age can be determined from (4.16) and (4.17), using (4.9), (4.25, (4.27), and (4.28).

Figure 4.5 displays the expected number of years in poor health between age 60 and retirement age, i.e.,

$$LE_{active}(t, g) - HLE_{active}(t, g),$$

for males (left panel) and females (right panel) of cohorts aged 60 in year $t = 2010, \ldots, 2045$.

The figure shows that the best-estimate expected number of years in poor health between age 60 and retirement age decreases slightly over time. There is, however, a significant degree of uncertainty. The upper bounds of the forecast intervals correspond to an increase in the expected number of years in poor health between age 60 and re-

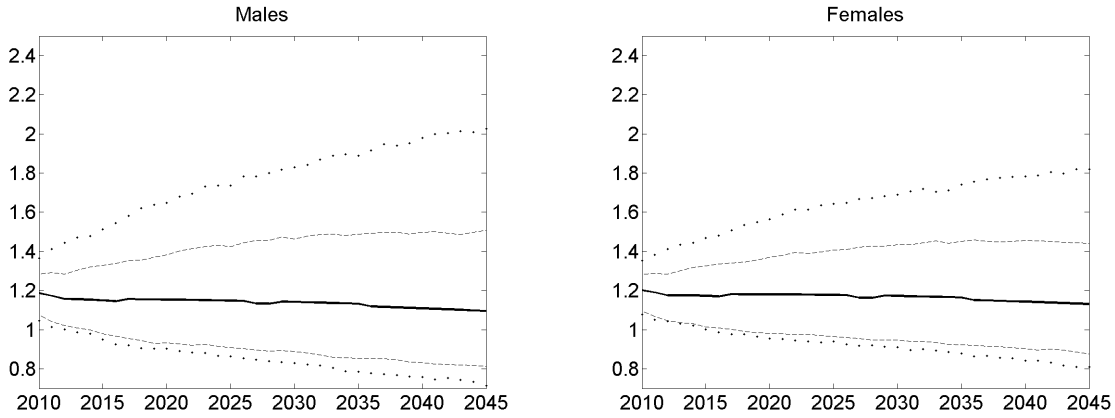

Figure 4.5 – Expected number of years in poor health between age 60 and retirement age, as a function of the year in which the individual reaches age 60, for males (left panel) and for females (right panel). In each figure, the solid line represents the best-estimate forecast; the dashed lines and the dotted lines represent the 95% forecast intervals, where the smaller intervals reflect process risk only while the wider intervals reflect both process risk and parameter risk.

tirement age from 1.4 years in 2010 to 2.0 years in 2045 for males. The corresponding numbers for females are 1.4 and 1.8. These results suggest that linking retirement age to life expectancy may lead to significant reductions in the fraction of individuals that remains sufficiently healthy to work until extended retirement age.

Health status at retirement age

In this section, we investigate the effect of the retirement age policy on health status at retirement age. We introduce the following notation:[10]

- $\varphi_{y,s}$: the probability of being alive at age y in year $2010+s$, conditional on being alive at age 30; the expression for $\varphi_{y,s}$ is presented in Appendix 4.A;[11]
- $\xi_{y,s}$: the probability of being in good health at age y in year $2010+s$, conditional on being alive at age y, i.e., $\xi_{y,s} = 1 - \pi^{(g)}_{y,2010+s}$.

Because retirement age of an individual with gender g belonging to the cohort aged x in 2010 is $R(x)$, the fraction of these individuals who survive until their retirement

[10] To avoid overloaded notation, we omit the gender index in some of our notation in this section.

[11] We condition on being alive at age 30 because of data availability. The cumulative survival probability between age x_0 and age $R(x)$ for the cohort aged x in 2010 depends on the one-year survival probability in years $2010 - x + x_0$ (the year in which they are x_0 years old) till $2010 - x + R(x)$ (the year in which they retire). Because we use data as off 1972, we need $2010 - x + x_0 \geq 1972$ for all $x \leq 66$, and, hence, $x_0 \geq 28$.

age, conditional on having been alive at age 30, is given by

$$P_{alive}(x,g) := \varphi_{R(x),R(x)-x},$$

and the corresponding fraction who is in good health is

$$P_{healthy}(x,g) = \varphi_{R(x),R(x)-x} \cdot \xi_{R(x),R(x)-x}.$$

Figure 4.6 displays $P_{alive}(x,g)$ and $P_{healthy}(x,g)$ for males (left panel) and females (right panel), respectively, as a function of the age x in 2010.[12] The results suggest that even though younger cohorts will retire at later ages than older cohorts, their probability to survive until retirement is higher. The effect is even stronger for the fraction that survives and is in good health at retirement age. This holds true for both males and females, although the effects for males are stronger than for females. For example, conditional on being alive at age 30, the fraction of individuals that survives until retirement age increases from 0.80 to 0.86 for male cohorts aged 66 till 25 in 2010; the fraction in good health at retirement age for these cohorts increases from 0.61 to 0.75. In general, mortality and health improve over time, but depreciate as people age. Our results suggest that the positive effects on mortality and health of improvements over time are stronger than the negative effects on mortality and health of delayed retirement.

Furthermore, Figure 4.6 also shows that, as was the case for (healthy) life expectancy after retirement, the difference in the likelihood of being in good health at retirement age between males and females is also projected to decrease. Whereas females are more likely to survive until retirement than males, and are also more likely to be in good health at retirement, these differences are projected to become smaller.

We conclude this section by investigating the effect of the policy on the likelihood of reaching retirement age in good health. Specifically, we let $\Delta_{healthy}(x,g)$ denote the difference between the fraction of people that is in good health at retirement when the policy is implemented, and what that fraction would be if retirement age would stay at 66 for all future cohorts, which we denote $P_{healthy|R(x)=66}(x,g)$. Then,

$$\Delta_{healthy}(x,g) = P_{healthy}(x,g) - P_{healthy|R(x)=66}(x,g) \tag{4.20}$$

$$= \varphi_{R(x),R(x)-x} \cdot \xi_{R(x),R(x)-x} - \varphi_{66,66-x} \cdot \xi_{66,66-x}. \tag{4.21}$$

Moreover, to gain insight into which factors contribute most to differences in survival in good health due to the policy, we decompose $\Delta_{healthy}(x,g)$ into a health effect (which

[12] Expected probabilities for cohorts who are between 60 and 66 years old are somewhat non-smooth over age. This occurs because for these cohorts, most of the data used to derive the cumulative survival probability as off age 30 are observed, rather than predicated.

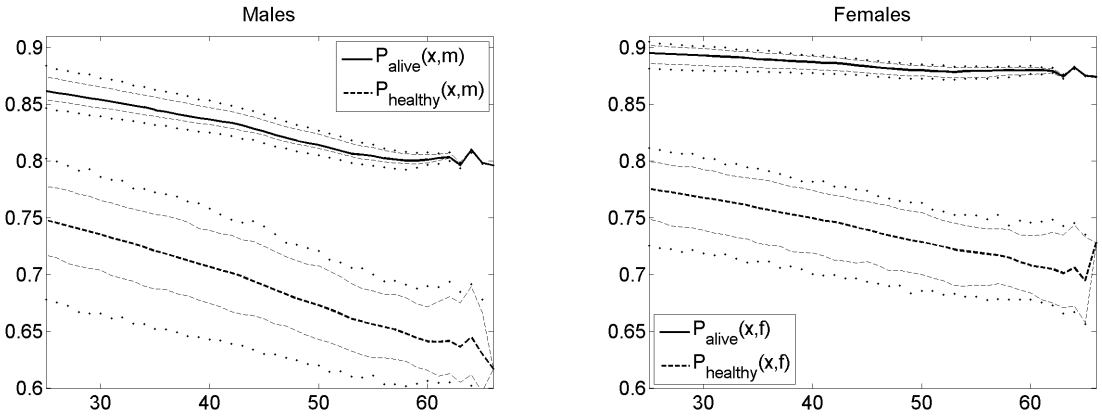

Figure 4.6 – Probability of being alive (upper lines) and probability of being alive and in good health (lower lines) at retirement age, conditional on being alive at age 30 for males (left figure) and females (right figure). In each figure, the darker solid lines and dahsed lines represent the best-estimate forecasts; the dashed lines and dottend lines represent the 95% forecast intervals, where the smaller intervals reflect process risk only while the wider intervals reflect both process risk and parameter risk.

we denote $H(x,g)$), and a mortality effect (which we denote $M(x,g)$), i.e.,

$$\Delta_{healthy}(x,g) = M(x,g) + H(x,g),$$

where

$$M(x,g) = \left(\varphi_{R(x),R(x)-x} - \varphi_{66,66-x}\right) \cdot \xi_{66,66-x}, \tag{4.22}$$

$$H(x,g) = \varphi_{R(x),R(x)-x} \cdot \left(\xi_{R(x),R(x)-x} - \xi_{66,66-x}\right). \tag{4.23}$$

The mortality effect $M(x,g)$ reflects the effect of changes in mortality on $\Delta_{healthy}(x,g)$ in case health prevalence at retirement age is the same as the health prevalence at age 66 ($\pi_{66,66-x}$). The health effect can be interpreted as the residual effect due to changes in health between age 66 and $R(x)$.

Figure 4.7 shows $\Delta_{healthy}(x,g)$ as well as the decomposition of the total effect into a mortality effect ($M(x,g)$) and a health effect ($H(x,g)$), see (4.21), (4.22), and (4.23), for males and females. As expected, as compared to the situation which the retirement age would stay at 66, the increase in retirement age leads to a smaller fraction that survives and is in good health at retirement age i.e., $\Delta_{healthy}(x,g)$ is negative for all cohorts and for both genders. However, the effect is very small. The best estimate projection suggests that over a period of 50 years, the fraction that survives and is in good health at retirement age would be less than 3 percentage points lower than when retirment

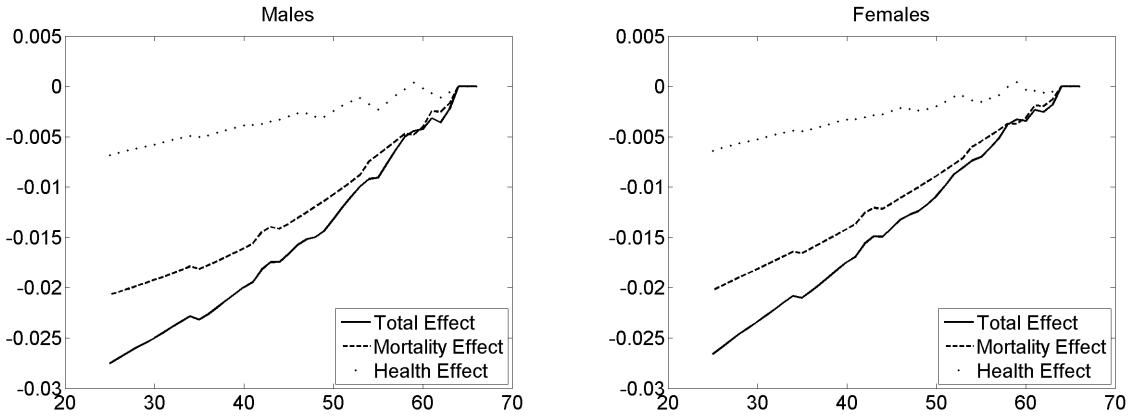

Figure 4.7 – Decomposition of the effect of the retirement age policy on the probability of being in good health at retirement age, for individuals aged $x = 25, \ldots, 66$ in 2010. The left panel corresponds to males; the right panel corresponds to females. The solid lines represent the total effect; the dashed lines represent the effect of increased mortality; the dotted lines represent the effect of decreased health.

age would not increase. Moreover, this effect is mainly driven by increased mortality due to the delay in retirement age. The effect of worsened health is negligible.

4.4 Conclusion

In this paper, we consider a policy in which, roughly speaking, an increase in gender-neutral life expectancy of one month is accompanied by an increase in retirement age of one month. In addition to avoiding that the number of years spent in retirement continues to increase as life expectancy increases, such policies have the important advantage that they can significantly reduce the longevity risk born by pension providers.

Our study investigates the effect of this policy on remaining (healthy) life expectancy before and after retirement, as well as on the likelihood of being in good health at retirement age. In addition to best-estimate projections, we quantify the degree of uncertainty using a stochastic forecast model that incorporates dependence between the developments of health and mortality over time. Based on best-estimate projections alone, one would be inclined to conclude that improvements in health of the elderly are sufficiently strong to justify linking retirement age to developments in life expectancy. However, the bounds of the forecast intervals that we generate correspond to an increase in the time spent in poor health between age 60 and retirement age by 2.5 months per decade for males, and by 1.6 months per decade for females. The bounds of the forecast intervals for the time spent in good health after retirement correspond to decreases by a little bit less than 5 months per decade for both males and females.

The extent to which our results allow to draw conclusions regarding the effect of the policy for work force participation depends on whether self-reported health helps to predict people's willingness and/or ability to work. Identifying a causal effect of self-reported health on retirement decisions is complicated due to potential endogeneity and justification biases in self-reported health (see, e.g., Bound, 1991; Bound, Schoenbaum, Stinebrickner, and Waidmann, 1999; McGarry, 2004; Lindeboom and Kerkhofs, 2009, to name just a few). Nevertheless, several studies provide evidence that individuals that report poor self-assessed health are more likely to retire early (see, e.g., McGarry, 2004; Jones, Rice, and Roberts, 2010; Au, Crossley, and Schellhorn, 2005).

We conclude by discussing potential directions for future reserach. First, it is well documented that, in addition to gender, (healthy) life expectancy and health status depend significantly on social status. This suggests that the effects of the policy that we have investigated can be significantly different for different social status groups (see, e.g., Cutler, Meara, and Richards-Shubik (2011) and Munnell, Meme, Jivan, and Cahill (2004)). Based on our current results, a preliminary conclusion could be that for individuals belonging to groups with weaker improvements in health and/or mortality than the overall population, a policy that links retirement age to population life expectancy could have rather strong adverse effects on their life expectancy and healthy life expectancy after retirement, and on the likelihood of being in good health at retirement age. Quantification of these effects is left for future research.

Second, a potential limitation of this study is that it ignores potential causal effects of retirement on mortality and health. Regarding the effects of retirement on health, it has often been argued that retirement negatively affects health due to, for example, declines in physical activity (see, e.g., Dhaval, Rashad, and Spasojevic, 2006). On the other hand, factors such as stress relief or more time to engage in healthy behavior after retirement could have the opposite effect (e.g., McGarry, 2004). While some studies find a negative effect of retirement on health (Dhaval, Rashad, and Spasojevic, 2006; Behncke, 2012), others find either no effect or a small positive effect (Bound and Waidmann, 2007; Neuman, 2008; Coe and Lindeboom, 2008; Pedersen and Bingley, 2011), or a significant but temporary positive effect (Coe and Zamarro, 2011). Moreover, some studies find different effects for different dimensions of health. van der Heide, van Rijn, Robroek, Burdorf, and Proper (2013) find a positive effect of retirement on mental health, and contradictory evidence of the effect of retirement on self-assessed health and physical health. Eibich (2013) finds a positive effect of retirement on subjective health and mental health. Regarding the effect of retirement on mortality, Hernaes, Markussen, Piggott, and Vestad (2013) find no effect. Given the inconclusive evidence regarding a causal effect of retirement on mortality and health, we have analyzed the effects of the retirement policy on healthy life expectancy before and after retirement, assuming no effect of retirement on mortality and health.

4.A Notation and formulas

In this appendix, we first introduce some notation, and give the expressions for period and cohort (healthy) life expectancy used in Section 4.3.2, as well as the cumulative survival probability used in Section 4.3.3.

4.A.1 Notation

We use the following notation:

- $\mathcal{M} = \{\frac{i}{12} : i = 0, \ldots, 12 * \omega\}$: the set of ages (expressed as multiples of months);
- ${}_{\frac{1}{12}}q_{x,t}^{(g)}$: the probability of death within the age interval $[x, x + \frac{1}{12})$, conditional on being alive at age x at time t, for ages $x \in \mathcal{M}\backslash\{\omega\}$ and gender group $g \in \{m, f, n\}$; we discuss the details of the estimation of these gender neutral death probabilities in Section 4.2.2; individuals who have reached age $\omega - 1/12$ are assumed to die within the age interval $(\omega - 1/12, \omega]$;
- ${}_sp_{x,t}^{period,(n)}$: the probability that an individual aged x in year t survives to age $x + s$, based on the *gender neutral period monthly lifetable* in year t. This survival probability is determined under the assumption that death rates for future years are identical to the death rates in year t, i.e., for any $x, s \in \mathcal{M}$ such that $x + s \leq \omega$, it holds that[13]
$$ {}_sp_{x,t}^{period,(n)} = \prod_{i \in \mathcal{M}: i < s} \left(1 - {}_{\frac{1}{12}}q_{x+i,t}^{(n)}\right); \tag{4.24}$$
- ${}_sp_{x,t}^{cohort,(g)}$: the gender-specific probability that an x-year old in year t survives to (at least) age $x + s$, based on the *monthly gender-specific cohort lifetable* in year t, i.e.,
$$ {}_sp_{x,t}^{cohort,(g)} = \prod_{i \in \mathcal{M}: i < s} \left(1 - {}_{\frac{1}{12}}q_{x+i,t+i}^{(g)}\right). \tag{4.25}$$

4.A.2 (Healthy) life expectancy

Gender neutral monthly period life expectancy of an x-year old in year t, which we denote $LE^{period}(x, t)$, can be determined as:

$$ LE^{period}(x, t) = \frac{1}{12}\left(\sum_{s \in \mathcal{M}: s < \omega - x} {}_sp_{x,t}^{period,(n)}\right). \tag{4.26}$$

The expected number of years lived between age x and age y for an x-year old in year t with gender $g \in \{m, f\}$, based on *gender-specific monthly cohort* probabilities can

[13] Formally, the assumption is that ${}_{\frac{1}{12}}q_{y,t+i} = {}_{\frac{1}{12}}q_{y,t}$ for all $i \geq 0$.

be determined as:

$$LE^{cohort}(x,y,g,t) = \frac{1}{12}\left(\sum_{s\in\mathcal{M}:s<y-x} {}_sp_{x,t}^{cohort,(g)}\right). \tag{4.27}$$

Moreover, following Imai and Soneji (2007b), the expected number of years lived in good health between age x and age y for an x-year old in year t with gender $g \in \{m, f\}$, based on *gender-specific monthly cohort* probabilities can be determined as:

$$HLE^{cohort}(x,y,g,t) = \frac{1}{12}\left(\sum_{s\in\mathcal{M}:s<y-x} (1-\pi_{x+s,t+s}^{(g)}) \cdot {}_sp_{x,t}^{cohort,(g)}\right). \tag{4.28}$$

4.A.3 Survival until retirement

Because an individual aged y in year $2010+s$ belongs to the cohort that was born in year $2010+s-y$, it holds that the expression for $\varphi_{y,s}$ used in Section 4.3.3 is given by:

$$\varphi_{y,s} = \prod_{i\in\mathcal{M}:30\leq i\leq y-1} (1 - {}_{\frac{1}{12}}q_{i,2010+s-y+i}^{(g)}). \tag{4.29}$$

4.B Parameter estimates of the mortality and health forecast model

Estimates of the Lee-Carter model with observed variables. Figure 4.8 show estimates for male and female mortality. Figure 4.9 for male and female health.

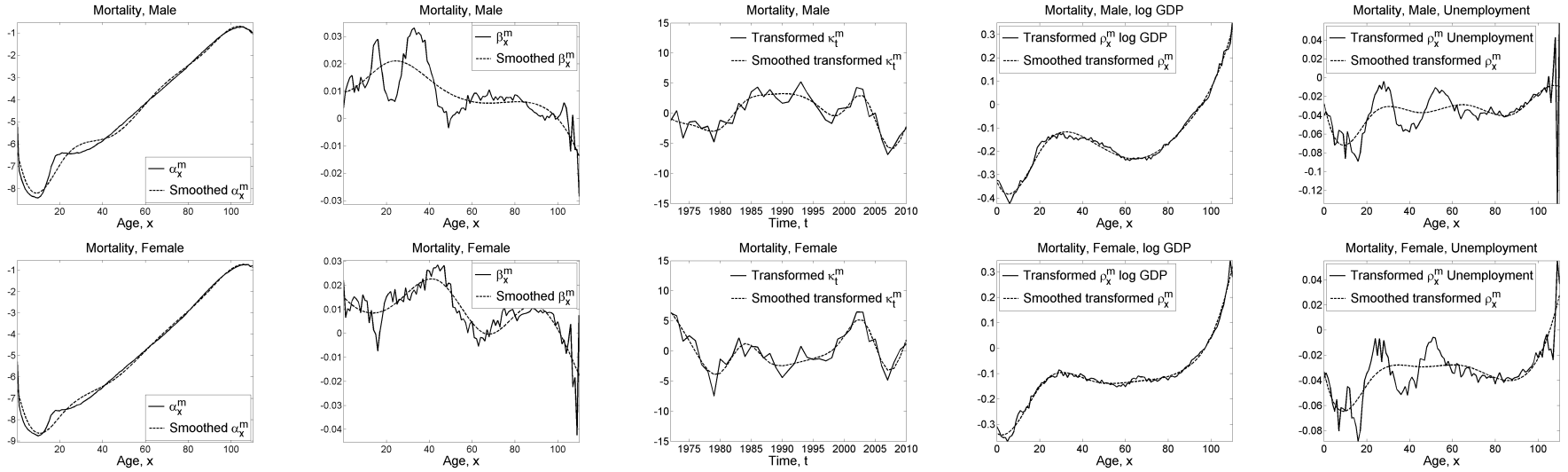

Figure 4.8 – Estimates of Lee-Carter model for mortality

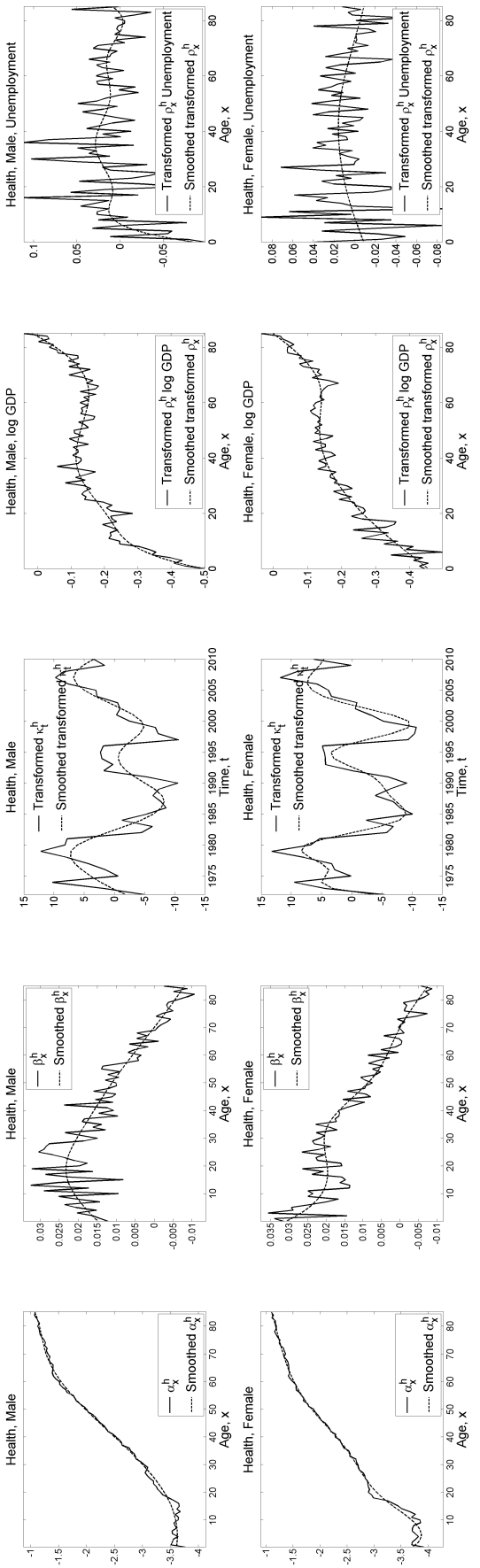

Figure 4.9 – Estimates of Lee-Carter model for health

Table 4.1 – Estimates of $VAR(3)$ model, equation (4.12)

$$\boldsymbol{Y}_t \equiv \begin{bmatrix}\Delta K_t \\ \Delta Z_t\end{bmatrix} = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \Theta_2 \boldsymbol{Y}_{t-2} + \Theta_3 \boldsymbol{Y}_{t-3} + \nu_t$$

$$\hat{\boldsymbol{C}} = \begin{pmatrix} -1.56(1.40) \\ -0.38(1.57) \\ -4.87(3.14) \\ -2.00(3.42) \\ -0.06(0.06) \\ 0.54(0.46) \end{pmatrix}$$

$$\begin{bmatrix}\Delta\kappa^m_{t-1,m} & \Delta\kappa^m_{t-1,f} & \Delta\kappa^h_{t-1,m} & \Delta\kappa^h_{t-1,f} & \Delta\log(GDP)_{t-1} & \Delta UnEmp_{t-1}\end{bmatrix}$$

$$\hat{\Theta}_1 = \begin{pmatrix} 0.01(0.52) & -0.08(0.43) & 0.06(0.17) & 0.02(0.16) & 15.04(10.53) & 2.27(1.66) \\ -0.16(0.58) & 0.02(0.49) & 0.18(0.19) & -0.03(0.18) & 12.16(11.84) & 1.86(1.87) \\ 2.25(1.17) & -2.17(0.97) & 0.12(0.37) & -0.25(0.36) & -66.57(23.62) & -9.39(3.73) \\ 2.97(1.27) & -2.84(1.06) & 0.90(0.40) & -0.94(0.39) & -76.33(25.74) & -10.43(4.06) \\ -0.06(0.02) & 0.05(0.02) & -0.02(0.01) & 0.01(0.01) & 1.02(0.47) & 0.09(0.07) \\ 0.26(0.17) & -0.21(0.14) & 0.10(0.05) & -0.04(0.05) & -4.95(3.50) & -0.09(0.55) \end{pmatrix}$$

$$\begin{bmatrix}\Delta\kappa^m_{t-2,m} & \Delta\kappa^m_{t-2,f} & \Delta\kappa^h_{t-2,m} & \Delta\kappa^h_{t-2,f} & \Delta\log(GDP)_{t-2} & \Delta UnEmp_{t-2}\end{bmatrix}$$

$$\hat{\Theta}_2 = \begin{pmatrix} -0.30(0.48) & 0.28(0.38) & 0.01(0.17) & -0.16(0.15) & 0.17(8.86) & -0.36(1.29) \\ -1.11(0.54) & 1.05(0.43) & 0.16(0.19) & -0.29(0.17) & 2.10(9.96) & -1.09(1.45) \\ -0.72(1.08) & -0.04(0.85) & 0.09(0.38) & 0.13(0.34) & 26.71(19.88) & 7.42(2.89) \\ 0.52(1.17) & -0.57(0.93) & 0.71(0.41) & -0.31(0.37) & 11.91(21.66) & 8.66(3.15) \\ 0.06(0.02) & -0.05(0.02) & -0.01(0.01) & 0.002(0.01) & -0.42(0.40) & 0.0002(0.06) \\ -0.37(0.16) & 0.30(0.13) & 0.002(0.06) & 0.02(0.05) & 4.69(2.94) & 0.03(0.43) \end{pmatrix}$$

$$\begin{bmatrix}\Delta\kappa^m_{t-3,m} & \Delta\kappa^m_{t-3,f} & \Delta\kappa^h_{t-3,m} & \Delta\kappa^h_{t-3,f} & \Delta\log(GDP)_{t-3} & \Delta UnEmp_{t-3}\end{bmatrix}$$

$$\hat{\Theta}_3 = \begin{pmatrix} 1.23(0.49) & -0.57(0.38) & 0.16(0.16) & -0.25(0.13) & -5.37(8.03) & 1.8097(0.93) \\ 1.83(0.55) & -0.95(0.43) & 0.31(0.18) & -0.35(0.15) & -12.79(9.03) & 1.1903(1.05) \\ -2.69(1.09) & 2.04(0.85) & -0.34(0.36) & 0.19(0.30) & 58.71(18.01) & 2.2562(2.10) \\ -3.03(1.19) & 2.82(0.93) & -0.15(0.39) & 0.01(0.33) & 73.32(19.63) & 3.9927(2.29) \\ -0.02(0.02) & 0.02(0.02) & -0.02(0.01) & 0.01(0.01) & 0.74(0.36) & 0.1497(0.04) \\ 0.10(0.16) & -0.11(0.13) & 0.07(0.05) & -0.05(0.04) & -5.53(2.67) & -1.0272(0.31) \end{pmatrix}$$

$$\hat{\Sigma} = \begin{pmatrix} 1.61 & 1.69 & 0.32 & 0.72 & 0.04 & -0.28 \\ 1.69 & 2.03 & 0.49 & 0.72 & 0.04 & -0.29 \\ 0.32 & 0.49 & 8.09 & 7.90 & 0.03 & -0.46 \\ 0.72 & 0.72 & 7.90 & 9.60 & 0.06 & -0.68 \\ 0.04 & 0.04 & 0.03 & 0.06 & 0.00 & -0.02 \\ -0.28 & -0.29 & -0.46 & -0.68 & -0.02 & 0.18 \end{pmatrix}$$

Note: The table presents estimates of $VAR(3)$ model for females with κ^m_m, κ^m_f, κ^h_m, κ^f, log GDP, and unemployment rate.
$\Delta K_t = (\Delta\kappa^m_{t,m}, \Delta\kappa^m_{t,f}, \Delta\kappa^h_{t,m}, \Delta\kappa^h_{t,f})'$, $\Delta Z_t = (\Delta\log(GDP)_t, \Delta UnEmp_t)'$, where $UnEmp$ denotes unemployment rate.
Standard errors are provided in parentheses.
$\hat{\Sigma}$ is the estimated variance covariance matrix of ν.

CHAPTER 5

AN ANALYSIS OF THE INTERACTION BETWEEN HEALTH EXPENDITURE AND ITS DETERMINANTS IN THE U.S.

This Chapter is based on Yang and Melenberg (2014)

This chapter investigates the dynamic economic relationship between healthcare expenditure, macroeconomic determinants, the age structure of the population, and the elderly's self-assessed health status. We investigate the dynamic relationship between health expenditure and its determinants after transformations to stationarity. We find that an improvement in the elderly's health status slows down the rising healthcare cost. The increase of the proportion of the elderly people in the population has a positive effect on the rising healthcare cost. Healthcare is found to be a necessity good after controlling for the other determinants. Moreover, relative healthcare price and public financing are significant factors affecting the increase of the healthcare cost. An out-of-sample prediction analysis shows that accounting for the elderly's health status helps to improve the accuracy of the health expenditure forecasts, compared to ignoring the relationship and to official (CMS) forecasts.

5.1 Introduction

In the United States, healthcare as a share of the country's GDP dramatically increased over the past half-century from 5.2% in 1960 to 17.9% in 2011[1]. This share as well as

[1]Source: "NHE summary including share of GDP, CY 1960-2011" provided by National Health Expenditure Data at Center of Medicare and Medicaid

the per capita spending on healthcare in the United Status are both much higher than in other countries. As is well documented by the current health economics literature, factors such as national income, economic growth, public financing of the healthcare, relative price of healthcare services, and age structure are main determinants of the rising health expenditure. In light of these findings, this paper proposes to investigate the dynamic relationship between health expenditure and a set of determinants in the United States, including additionally the elderly's health status, a potentially important determinant of the demand for healthcare.

Health expenditure and the determinants that we consider all show a clear trending behavior. However, our test results suggest that these variables show different nonstationary behavior, including trend stationarity, first difference stationarity, and second difference stationarity processes. These different forms of nonstationarity complicate the econometric analysis considerably, because their presence prevents the straightforward use of many standard econometric models. In this paper we proceed as follows to deal with these various forms of nonstationarity. First, we apply variable-specific transformations so that after applying these transformations the transformed variables are (close to) stationary. In particular, we find that the logarithm of health expenditure is integrated of order one, so that its corresponding first difference (the growth rate in health expenditure) is stationary. Then we apply a Vector Auto Regression (VAR) model to capture the joint dynamic relationships between these stationarized variables. This means that we model the growth rate in health expenditure jointly with the other stationarized variables, allowing the other variables to influence health expenditure but also the other way around. By including a relatively large number of variables, we might be able to reduce a potential bias due to omitted variables. The VAR model can straightforwardly be used to make forecasts for all variables included, in particular, the growth rate in health expenditure. We investigate the out-of-sample forecasting performance of our VAR model and make a comparison with "official forecasts."

As discussed in the next section, there is some controversy in the literature whether health care is a luxury or a necessity good and whether the relative healthcare service price and the public financing of the healthcare sector have a positive or negative effect of healthcare spending. We find that the income elasticity is less than one, so that according to our estimates healthcare is a necessity good and our results show that both the stationarized relative healthcare service price and the stationarized ratio of public to total financing of the healthcare sector have a positive effect on healthcare spending. We also investigate the importance of "Baumol's cost disease," and find that the stationarized variable to quantify this "disease" turns out to have only a small impact. On the other hand, productivity growth shows a clear positive effect. Furthermore, both the proportion of the younger and the older elderly have a positive effect on health expenditure (after stationarization). Finally, we find that an improvement in the (stationarized) elderly's health status implies a slower growth of the per capita healthcare

spending. Our out-of-sample forecasts show that, although forecasting during the financial crisis is hard, the forecasts based on the VAR-model might be useful to help to improve the official forecasts.

The remainder of this paper first discusses the current health economics literature on explaining the main factors account for the increase in the healthcare spending and the literature on the methodology of analyzing this growth. We then describe the variables used in this analysis and the available data in Section 5.3. The Appendix contains further information which transformations we used to stationarize our variables. Section 5.4 describes the VAR model applied in this analysis and presents the corresponding estimation results. We especially focus on the effect of the elderly's health status on the growth of the health expenditure. Forecasts of future health expenditure are also presented. We conclude in Section 5.5.

5.2 Overview of the literature

The analysis of drivers behind the evolution in healthcare spending is far from straightforward. It should take account of a complex network of interactions between factors from the demand side and the supply side. This section first is going to summarize findings from various studies for a number of possible variables on determining healthcare expenditure from both the demand and the supply sides. Then we provide a brief overview of the current methodology applied in studying the rising health expenditure.

5.2.1 Drivers of healthcare spending - factors that affect demand

Income—The growth of the national income has been recognized as an important determinant in the growth of healthcare expenditure by many studies, including both cross-country studies and single-country studies. In an early study, Newhouse (1977) has proposed that national income can explain much of the increased health expenditure. Hall and Jones (2007) also suggest that the growth of health spending is a rational response to the growth of income per person. Similar conclusions are derived by other researchers, such as Gerdtham, Jonsson, MacFarlan, and Oxley (1998), Roberts (1999), Oliveria Martins, De la Maisonneuve, and Bjornerud (2006), Christiansen, Bech, and Lauridsen (2007), and Amiri and Ventelou (2012).

Many studies have been focusing on the estimation of the income elasticity of health expenditure, questioning whether healthcare is a luxury or a necessity good (see Hansen and King (1996), Blomqvist and Carter (1997), Gerdtham and Lothgren (2000), Dreger and Reimers (2005), Baltagi and Moscone (2010), and Moscone and Tosetti (2010), etc.). However, the complexity of estimating the income elasticity can be attributed to at least two reasons. First, the estimation might be biased due to an omitted variable bias, see

Roberts (1999) and Gerdtham and Jonsson (2000). For example, omitting the influence of healthcare service price or population ageing may overestimate the income elasticity. Second, there might exist a simultaneous relationship between income and health expenditure. Failure to take into account the impact of the healthcare spending on national income may also lead to a biased estimation of income elasticity, see Xu, Saksena, and Holly (2011).

Demographic structure—Population ageing is increasingly identified as an important factor affecting the rising healthcare spending. The importance of ageing on the rising health expenditure are both examined by studies concerning the United States (see for example Murthy and Ukpolo (1994) and Murthy and Okunade (2000)), and concerning other OECD countries (see for example Dreger and Reimers (2005) and Okunade, Karakus, and Okeke (2004)). However, researchers including Zweifel, Felder, and Meiers (1999) and Yang, Norton, and Stearns (2003) point out that because the largest healthcare expenditure typically happens during the last years of one's life, the impact of population ageing on healthcare expenditure may be estimated with a bias if the time remaining to death is not taken into account. For this reason, Okunade, Karakus, and Okeke (2004) suggest that "healthcare costs are better off shifting demographic paradigms from measurements based on simple population aging to that of 'time to death'" (Page 175). Researchers such as Zweifel, Felder, and Werblow (2004) and Werblow, Felder, and Zweifel (2007) empirically tested the positive effect of the variable "time to death" on the rising healthcare cost. When analyzing the effect of the population ageing on the growth of the healthcare cost, instead of including the "time to death" variable, some researchers also separate the elderly age group into the younger elderly and the older elderly. For example, Christiansen, Bech, and Lauridsen (2007) separately estimate the effects of the population groups aged 65-74 and aged 75+ on the health expenditure. They find that these two groups have different effects. van Elk, Mot, and Franses (2009) apply a panel data approach and find that both the ratio of population aged 65-74 and aged above 75 to the total population have a significant positive effect on the rising health expenditure.

Health status of the elderly— The increase in the share of the older people in the population has an obvious impact on the demand for healthcare. However, as the demand for healthcare is ultimately derived from being in good health or not, age itself may not be a sufficient factor to explain the rising health expenditure. For this reason, the elderly's health status has been considered in analyzing healthcare spending. In many studies, researchers usually use the "elderly's health" and the "demographic structure" interchangeably. For example, Leu (1986), Cutler and Sheiner (1998), and Colombier (2012) use the percentage of people over 65 in the population as a proxy for population health. Lee and Miller (2002) adopt "time to death" as a rough indicator for the elderly's health status and project the future health expenditure by assuming a

fixed schedule relating health expenditure to "time to death." Other proxies, such as life expectancy, death rate, and disability, are also commonly applied as representing the health status, see, for example, Dormont, Grignon, and Huber (2006), Manton, Lamb, and Gu (2007) and Colombier (2012). Recently, Solakoglu and Civan (2012) use the population self-reported health as an indicator of healthcare need and find that the increased need for healthcare can explain the rising share of healthcare expenditure in GDP for the studied OECD countries.[2]

Relative price of the health service to the GDP deflator— The relevance of including the relative price of healthcare is addressed in van Elk, Mot, and Franses (2009). In studies of OECD countries, they obtain large positive effects of the relative price in the short run. Okunade, Karakus, and Okeke (2004) find that the increase in healthcare price significantly drives up the healthcare costs during different periods. Roberts (1999) finds that, in the long run, the effect of the relative price falls from positive to negative, during the examined period 1960 to 1993. Murthy and Ukpolo (1994) conduct a cointegration study for the aggregated health expenditure for the United States, obtaining a long-run negative effect.

5.2.2 Drivers of healthcare spending - factors that affect supply

In addition of factors on the demand side of the healthcare, factors on the supply side also affect the level of healthcare spending.

Baumol's cost disease—Baumol (1967) proposes the so called "Baumol's cost disease" to explain the growth of healthcare spending. In his study, he distinguishes two economic sectors, namely a "progressive" and a "nonprogressive" sector. Baumol (1967) characterizes the progressive sector, such as manufacturing sector, as being more capital intensive with a high level of productivity growth. Contrarily, nonprogressive industries, such as medical care and education, are seen as being more labor intensive with a relatively slow productivity growth. Because of higher productivity growth, the wage rate in the progressive sector tends to increase. If the ratio of the outputs of the two sectors is held constant, there has to be shifts of employment from the progressive sector to the nonprogressive sector. To do so, the nonprogressive sector has to increase the wage rate to attract more workers. As a result, the labor costs in the nonprogressive industries tend to increase because of an increase in wage rates in the progressive sector. In turn, the total costs in the nonprogressive sector, such as the healthcare sector, rises.

[2]In support of using self-reported health as a proxy for the health status, it is worth mentioning that McGee, Liao, Cao, and Cooper (1999), Hillen, Schaub, Hiestermann, Kirschner, and Robra (2000), and Burstrom and Fredlund (2001) find a positive relationship between self-reported and actual health outcomes, and empirically provide evidence of self-reported health to be a reasonable indicator for health.

One problem when estimating the effect of "Baumol's cost disease" is the complexity of measuring output and price deflators (Hartwig (2011) and Colombier (2012)). Without a proper measurement, testing whether "Baumol's cost disease" applies to the healthcare sector might be problematic. Recently, Hartwig (2008) proposed a new "Baumol's variable" to avoid the drawback of using medical-price indices. It is defined as the excess wage growth above the productivity growth, and equals the growth in unit costs in the Baumol sector. Applying this "Baumol's variable", helps to avoid a possible bias due to the use of an incorrect medical price deflator. Hartwig (2008) still obtained the conclusion that the healthcare costs' growth over the past 40 years in OECD countries can be attributed to the "Baumol's cost disease". In recent studies, Colombier (2012) and Bates and Santerre (2013) adjust "Baumol's variable" by the proportion of the labor force in the healthcare sector (or Baumol's sector) in the total economy. They show that after correcting the proportion of the labor participation, "Baumol's cost disease" is still valid in the healthcare. However, they obtained a smaller effect than the one provided by Hartwig (2008).

Public financing of the healthcare— A large part of the healthcare is financed by public funding. Therefore, the ratio of the public healthcare expenditure to the total healthcare expenditure (public financing of the healthcare) may be correlated with changes in the health expenditure. Researchers hold two different views on the role of public funding on the rising healthcare cost. Leu (1986) and his proponents, such as Gerdtham, Sogaard, Andersson, and Jonsson (1992), Murthy and Ukpolo (1994), and Murthy and Okunade (2000), support the view that public financing of the healthcare increases the total health expenditure. Leu (1986) explains this positive effect from two aspects. First, compared with the private sector, because of less competition in the public sector, the incentives to minimize costs might be lower. Moreover, bureaucrats might have incentives to maximize the government budget. Second, health insurance provided by the government reduces the healthcare price to consumers. Therefore, individuals tend to overuse healthcare services. On the contrary, Culyer (1989), Okunade, Karakus, and Okeke (2004), and others hold an opposite opinion. Okunade, Karakus, and Okeke (2004) find that "systems with greater government involvements in healthcare provision and financing can moderate spending growth" (page 179) for different periods among OECD countries. Gerdtham and Jonsson (2000) summarize a few reasons for this. Those are higher selling and advertising cost, less reliable market pressure, and higher production cost in the private sector.

Medical technology— Technology advance is considered to have a complicated impact on changes in healthcare costs. In general, it affects health expenditure from two directions. On the one hand, it increases the healthcare spending because of the invention of new and expensive medical technologies, see Zweifel (1984), Berndt, Cutler, Frank, Griliches, Newhouse, and Triplett (2000), Gerdtham and Lothgren (2000), and

Okunade (2001). Moreover, better technology prolongs the length of ill lives, but may result in higher spending on curing chronic diseases for a longer period, see van Elk, Mot, and Franses (2009). On the other hand, it reduces the healthcare costs because of several reasons. First, the availability of more advanced and better technology may induce more outpatients and, in turn, reduce cost for inpatient hospital stays. Second, new treatments may improve people's health status. As a result, the demand for healthcare will decline, see Hall and Jones (2007).

Measuring technology progress is a challenging issue in the current literature. Proxies are typically used, for example, health related R&D or total R&D (Okunade and Murthy (2002)), life expectancy, infant mortality, the share of the elderly (Dreger and Reimers (2005)), and diagnosis procedures such as MRI (Chandra and Skinner (2011)). Moreover, some researchers implicitly quantify the medical technology as a residual effect, after taking into account other factors in the equation for healthcare costs, see, for example, Newhouse (1992) and Oliveria Martins, De la Maisonneuve, and Bjornerud (2006). Others include a constant or a simple linear trend to capture in general terms technology progress, see for example Di Matteo (2005).

5.2.3 Overview of approaches studying healthcare expenditure

Different models have been used to study the link between healthcare spending and its determinants. Gerdtham and Jonsson (2000), Moscone and Tosetti (2010), and Xu, Saksena, and Holly (2011) provide reviews of current approaches applied to OECD countries and the United States. Depending on the type of data used, the main methods used in analyzing the determinants of health expenditure are cross-section analyses (Newhouse (1977), Leu (1986), Gerdtham, Sogaard, Andersson, and Jonsson (1992)), panel data analyses(Gerdtham (1992), Roberts (1999), Gerdtham and Lothgren (2000), and Xu, Saksena, and Holly (2011)), and unit root and cointegration analyses in a time series context (Murthy and Ukpolo (1994), Hansen and King (1996), King and Hansen (1996), and Blomqvist and Carter (1997)). Moreover, the unit root and cointegration methods are applied in analyzing not only single country aggregate healthcare expenditure (see Murthy and Ukpolo (1994), King and Hansen (1996), Murthy and Okunade (2000), Okunade and Murthy (2002), and Murthy (2012) for the studies in the U.S., see Hansen and King (1996), and Blomqvist and Carter (1997) for country-by-country studies for OECD countries), but also cross countries/states' health changes in a panel data framework (see Wang and Rettenmaier (2007) and Moscone and Tosetti (2010) for U.S. studies, and see MacDonald and Hopkins (2002), Dreger and Reimers (2005), and Baltagi and Moscone (2010) for OECD countries' studies).

If the time series are all integrated of the same order, and a linear combination of them is stationary, we say that these time series are cointegrated. Therefore, testing the stationarity of the studied time series has been addressed as a very important issue, be-

fore applying regression methods, when looking at the link between healthcare spending and its determinants at the aggregate level, see Gerdtham and Lothgren (2000), Moscone and Tosetti (2010), and Xu, Saksena, and Holly (2011). It is well known that the violation of the assumption that all time series are stationary might lead to spurious statistical results using OLS estimation (Engle and Granger (1987)). Appropriately choosing a model specification (whether to include a time trend or a constant) is an empirically difficult task for researchers when testing the existence of the unit root in the time series. Therefore, we observe conflicting conclusions regarding the stationarity/non-stationarity of health expenditure and its determinants, and possible cointegration relationships between them, see also reviews provided by Gerdtham and Lothgren (2000) and Moscone and Tosetti (2010).

Alternatively, if health expenditure and its determinants are not all integrated of the same order, a vector autoregression (VAR) model can be estimated, applied to the stationarized variables, to investigate the simultaneous relationship. In the area of health expenditure analysis, Erdil and Yetkiner (2009), Hartwig (2010), and Amiri and Ventelou (2012) are examples of cross-country studies. One potential weakness of the current research applying VAR models is that typically only the relationship between health expenditure, economic growth, and GDP is studied, while other potentially important determinants are omitted.

5.3 Data

In this section we first present the variables to be used in this paper. We then describe the data sources to obtain our time series data, which consists of the years 1972 to 2010 at an annual frequency. All our variables show a trending behaviour over this sample period. Therefore, we conclude this section by selecting appropriate data transformations to transform our variables into more stationary ones.

This study aims to consider a comprehensive set of determinants when analyzing the rising healthcare expenditure in real term. We consider eight variables as suggested by previous studies which may determine the increasing healthcare cost. The combined nine variables of interest will be denoted as follows,

- $PHE =$ Total health expenditure per capita.
- $INCOME =$ Personal income per capita.
- $RPHC =$ The ratio of healthcare service price index to GDP deflator with base year 2005.
- $AdjBV =$ Adjusted Baumol's variable.
- $PF =$ The ratio of public health expenditure to total health expenditure.

- $PROD$ = Labor productivity per person employed.
- AGE_{65-84} = The ratio of population 65 to 84 years old to total population.
- AGE_{85+} = The ratio of population 85 years and over to total population.
- HSG_{65+} = The share of the population who report good health within the age group 65 and older to the total population.

The National Health Expenditures Accounts (NHEA) released in 2011[3] provides annual total health spending per capita (PHE) in the United States. These are the official estimates of the Center of Medicare and Medicaid (CMS) database. The OECD (Organization for Economic Cooperation and Development) Statistics[4] provide the ratio of public healthcare expenditure to total healthcare expenditure (PF), the labour productivity per person employed ($PROD$), and the labour compensation per employee ($WAGE$), where the latter will be used to construct Baumol's variable. The OECD statistics define $PROD$ as real output (gross value added) divided by total employed persons, and define $WAGE$ as compensation of employees[5] divided by total employees[6]. The Bureau of Economic Analysis (BEA)[7] produces economic accounts statistics for personal income ($INCOME$). It is defined as the income received (including compensation and interest and dividend income) by persons from participation in production and from transfers from government and businesses. Personal income is closely monitored both as an indicator of economic activity and as a predictor of future spending. BEA also provides the price index of GDP, and the price index of healthcare services, from which we can derive the relative price of healthcare ($RPHC$). Total health expenditure per capita is deflated by the price index of healthcare services (base year 2005). All other economic variables are measured in U.S. dollar constant prices (OECD base year 2005).

The adjusted Baumol's variable ($AdjBV$) is calculated by the method provided and applied in Hartwig (2008) and Hartwig (2011). That is

$$AdjBV = \frac{1}{l_H}(\log(WAGE) - \log(PROD)), \tag{5.1}$$

[3]See http://www.cms.gov; go to "Research, Statistics, Data & Systems," then go to "National Health Expenditure Data," and finally go to "Historical."

[4]See http://stats.oecd.org/.

[5]Compensation of employees in the OCED dataset is defined as the sum of wage rates, earnings, employer contribution to statutory social security schemes or to private funded social insurance schemes, and unfunded employee social benefits paid by employers.

[6]The total number of employment and the total number of employees documented by the OECD dataset are classified from the United Nations International Standard Industrial Classification (ISIC) of all Economic Activities. This classification is the international standard for the classification of productive economic activities.

[7]See http://www.bea.gov/itable/index.cfm.

where l_H is the ratio of the employment in the healthcare sector to the total economy. To construct the adjusted Baumol's variable ($AdjBV$), we obtain detailed employment information from the Bureau of Economic Analysis (BEA). On October 5, 2006, BEA released estimates of full-time and part-time employment by industry for 1948-1997, based on the 1997 North American Industry Classification System (NAICS). Together with the data released for the years 1998-2012, we obtain the total number of employment in the economy and the total number of employment in the healthcare service. The ratio of the employment in the healthcare sector to the total economy (l_H) in turn can be derived. The variable $AdjBV$ can then be computed from equation (5.1).

Since the largest share of the health expenditure happens at the end of one's lifetime, the increased life expectancy has two effects. First, it means that a smaller share of the younger elderly will be in the last year of life. Furthermore, more of the elderly will be dying at older ages and the share of elderly's death shifts from the younger to the oldest old. The end-of-life expenditure will get pushed into the future. So, Stearns and Norton (2004) conclude that omitting the factor "time to death" will result in a biased estimation of the future health expenditure. Therefore, we separate the elderly age structure of the population into two age groups, namely the younger elderly who are between age 65 and 84 (AGE_{65-84}), and the oldest old whose age are 85 and over (AGE_{85+}). In this way, we approximately take into consideration the effect of "time to death" by the oldest old age group. The age-specific population data, which can be used to construct the age structure are obtained from the Human Mortality Database (HMD)[8].

Finally, we focus on the health status of the people 65 and over. This variable is different from Solakoglu and Civan (2012), who use the health status of the total population to indicate the healthcare demand. Because the elderly is the group which requires the majority of the healthcare, we use the elderly's health condition instead of the health status of the total population. The elderly's health status is obtained from the Integrated Health Interview Series (IHIS).[9] The National Health Interview Survey (NHIS) conducts annual surveys for civilian, non-institutionalized U.S. population. One of the survey questions asks respondents to rate their health with four categories ("excellent", "good", "fair", or "poor") before 1982 and five categories ("excellent", "very good", "good", "fair", or "poor") after 1982. The IHIS documents the integrated self-assessed health from survey results and provides annual data to track health status at the national level. Following Yang, De Waegenaere, and Melenberg (2013b), the health status variable of age 65 and over (HSG_{65+}) in this paper is measured as the proportion of the population of age 65 and over who report their health

[8]See http://www.mortality.org/

[9]For general information see `http://www.ihis.us/ihis/`. The health data was downloaded via the variable "Health status", available from the website `https://www.ihis.us/ihis-action/variables/group/health_general`.

status with "excellent", "very good," or "good" to the total surveyed population. Due to the data availability of the health status, this paper uses data from 1972 to 2010.

In Figure 5.1 we present the time series characteristics of our variables. All our variables show a clear trending behaviour over the sample period. Therefore, we select for each of our variables an appropriate transformation, so that after transforming the variable it becomes stationary. The details of this analysis are presented in the appendix. It turns out that different variables require different transformations. We start from the logarithmic form for all variables except for the adjusted Baumol's variable. We find that $\log(PHE_t)$, $\log(INCOME_t)$, $\log(PF_t)$, and $\log(PROD_t)$ are $I(1)$, requiring time differencing to make these variables stationary. We find that $\log(RPHC_t)$ and $\log(Age_{65-84,t})$ are $I(2)$, requiring double time differencing to make these variables stationary. Finally, we find that the remaining variables ($AdjBV_t$, $Age_{85+,t}$, and $HSG_{65+,t}$) are trend stationary, requiring detrending to make these variables stationary. The stationarized variables will be denoted by adding a superscript s. Thus, for example, PHE_t^s stands for $\Delta \log(PHE_t)$.

5.4 VAR models and empirical results

In the analysis of health expenditure and its determinants, it is very likely that there are bilateral relationships. For example, there is possibly a bilateral relationship between health expenditure and the elderly's health condition. On the one hand, an increase in the fraction of elderly in good health may slow down the increase in healthcare cost; on the other hand, an increase in healthcare expenditure may be devoted to provide a better medical treatment, improving the quality of life and, in turn, might help to improve people's health. Next, income is likely to affect the health expenditure, but a reverse effect may exist as well. Erdil and Yetkiner (2009) explained this reverse causation using two arguments. First, a higher health expenditure might positively affect human capital through the enhancement of education, which may lead to productivity growth and ultimately an increase in national income. Second, a better health condition of people may be achieved because of higher health expenditure. This may improve the labor participation and productivity growth[10]. In addition, it also seems reasonable to expect that even if causality exists in both directions, it does not occur instantaneously, but with some time lag. For these reasons, a VAR model is used in this analysis.[11]

[10] Weil (2007) quantitatively calculate the income gain because of health improvement. The paper states that "healthier people are better workers. They can work harder and longer and also think more clearly" (Page 1266). Bloom and Canning (2005) also find a positive relationship between health and productivity.

[11] The vector autoregressive (VAR) model has been very popularized by Sims (1980).

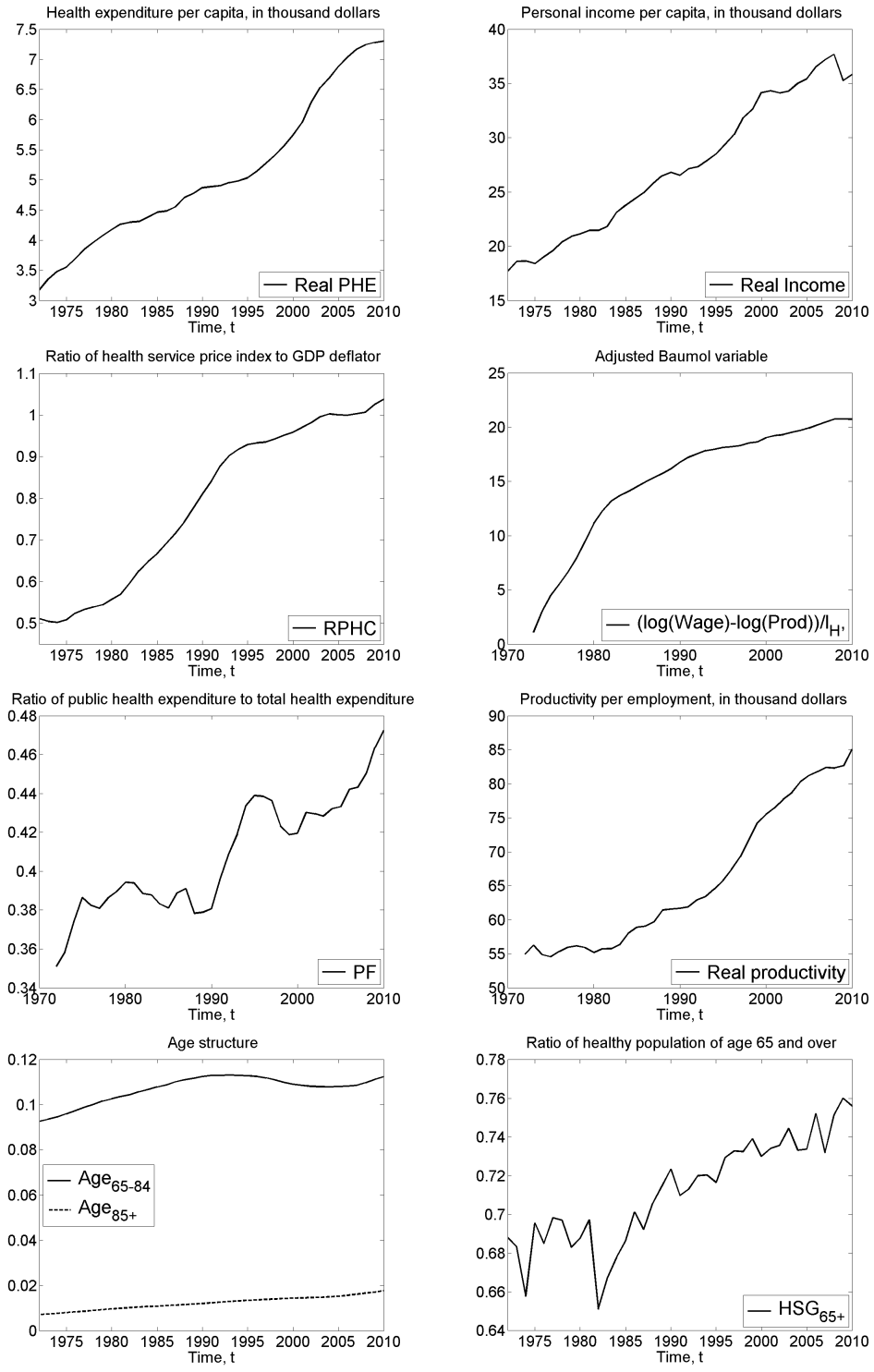

Figure 5.1 – Time series plots of the variables used in the analysis. The main text contains the definitions of these variables as well as the data sources.

A VAR model with p lags for the K-dimensional vector $\boldsymbol{Y}_t$ is specified as

$$\boldsymbol{Y}_t = \boldsymbol{C} + \sum_{i=1}^{p} \Theta_i \boldsymbol{Y}_{t-p} + \nu_t, \tag{5.2}$$

where Θ_i is a $K \times K$ coefficient matrix, $\boldsymbol{C}$ is a $(K \times 1)$ vector of intercepts allowing for possibility of a nonzero mean of $E(\boldsymbol{Y}_t)$. Finally, ν_t is a K-dimensional vector of white noise terms with zero mean and covariance matrix Σ_ν. In this study, $\boldsymbol{Y}_t$ is a vector containing the stationarized health expenditure and its stationarized determinants, i.e., $\boldsymbol{Y}_t$ is given by

$$\begin{aligned} \boldsymbol{Y}_t \;=\; & [PHE^s_t, INCOME^s_t, RPHC^s_t, AdjBV^s_t, PF^s_t, PROD^s_t, Age^s_{65-84,t}, \\ & Age^s_{85+,t}, HSG^s_{65+,t}]' \end{aligned} \tag{5.3}$$

The number of lags p in the VAR model can be selected by different information criteria, such as the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The information criteria select a model which has an optimal trade-off between model fit and parsimony, quantified in terms of number of parameters. Based on the loglikelihood function values (LLF) and the number of parameters (N_p), AIC and BIC are computed as follows

$$AIC(p) = (-2LLF + 2N_p) \tag{5.4}$$

$$BIC(p) = (-2LLF + 2N_p \log T) \tag{5.5}$$

For more information on the use of model selection criteria in VAR models, see chapter four in Lutkepohl (2007). Both AIC and BIC indicate a lag length of 1 ($p = 1$) in (5.2). Therefore, we estimate a VAR(1).[12]

As the elderly's health status can be an important indicator for healthcare spending, and the adjustment of health expenditure to the changes of elderly's health might even have lagged effect, we further examine the effect of elderly's health status up to the previous five years, i.e., we consider the VARX model

$$\boldsymbol{Y}_t = \boldsymbol{C} + \sum_{i=1}^{p} \Theta_i \boldsymbol{Y}_{t-p} + \sum_{i=2}^{5} \Phi_i HSG^s_{65+,t-i} + \nu_t, \tag{5.6}$$

where $HSG^s_{65+,t-i}$ are the extra "X"-variables, with corresponding vectors of regression coefficients Φ_i, $i = 2, \cdots, 5$. By assumption, these exogenous variables collected in X have a unidirectional effect on $\boldsymbol{Y}$. Both AIC and BIC also in this case indicate a

[12] To check the stability of this VAR(1), we check the inverse roots of the estimated characteristic AR polynomial (see Lutkepohl (2007)). The estimated VAR is stable if all roots have modulus less than one and lie inside the unit circle. The estimated VAR has as largest modulus 0.87, and all roots lie inside the unit circle. Therefore, our estimated VAR satisfies the conditions for stability.

lag length of 1 ($p = 1$).[13] To see the joint significance of the lagged elderly's health status included, F-tests are performed, suggesting that only the longest lag seems to be significant.[14] Therefore, we also investigate a VARX model with only $HSG^s_{65+,t-5}$ as extra "X"-variable. Both AIC and BIC again indicate a lag length of 1.[15]

Estimation Results—As the purpose of this study is investigating how health expenditure is affected by its determinants, we focus on the equation in the VAR model with PHE^s_t as the dependent variable. The corresponding estimation results, reflecting the short term effects, are shown in Table 5.1. The full estimation results are presented in Tables 5.12–5.14. We also look at the medium term cumulative effect on the health expenditure. Figure 5.2 shows 10 years cumulative impulse responses due to one unit change in each one of the error terms (ν_t). For example, assume that the error term in the equation of $INCOME^s_t$ increases by one unit, the upper left graph shows the cumulative changes in $\log(PHE_t)$ (not stationarized) after 1, 2, and so on, up to 10 years.

We obtain an income elasticity less than one, thus supporting the hypothesis that healthcare is a necessity good. Our estimated income elasticity is around 0.29 to 0.32 after three to four years. This is slightly less than Moscone and Tosetti (2010) who report an income elasticity of 0.36, which is at the lower end typically found in the literature.[16] Our relatively low income elasticity may due to the fact that we consider a relatively large set of determinants besides income, such as, for example, productivity, public finance, age structure, or the health status. Without these variables, income would capture their effects.

Furthermore, we find that the (stationarized) relative healthcare price has a clear positive effect on the health expenditure. Also public financing (after stationarization) has a clear positive effect on healthcare spending. This finding is in line with findings in Leu (1986), Murthy and Ukpolo (1994), and Murthy and Okunade (2000), supporting the view that provision of health care is less efficient in the public than in private sector. For the (stationarized) adjusted Baumol's variable we find a significant negative, but very small, short term effect, which is turned into a small positive effect in the medium term (after two to five years). Productivity per employment (after stationarization) does not seem to have a significant short term effect on the healthcare costs in the models with higher order terms of the lagged health status included. However, when turning to the medium term, also in the VARX-models the productivity per employment shows a clear positive effect on healthcare costs. Productivity growth

[13] By checking inverse roots of the characteristic AR polynomial, we find this VARX model stable.

[14] F-statistics are 4.225 with a p-value of 0.006 for $H_0 : \phi_{21} = \phi_{31} = \phi_{41} = \phi_{51} = 0$, and 0.678 with a p-value of 0.576 for $H_0 : \phi_{21} = \phi_{31} = \phi_{41} = 0$.

[15] Tests on the inverse roots of the characteristic AR polynomial indicates that the VARX model is stable.

[16] For example, Wang and Rettenmaier (2007) obtain as lowest income elasticity among 48 studied states a value of 0.514.

Table 5.1 – Estimates of equations with PHE_t^s as the dependent variables in the VAR model

	VAR Model	VARX Model 1	VARX Model 2
PHE_{t-1}^s	0.6602**(0.1900)	0.9578**(0.1903)	0.9953**(0.1774)
$INCOME_{t-1}^s$	0.2171**(0.0912)	0.2044**(0.0770)	0.2108**(0.0741)
$RPHC_{t-1}^s$	0.5878**(0.1971)	0.4737**(0.1834)	0.4665**(0.1684)
$AdjBV_{t-1}^s$	-0.0034 (0.0039)	-0.0098**(0.0044)	-0.0087**(0.0035)
PF_{t-1}^s	0.3489**(0.1643)	0.3910**(0.1512)	0.4388**(0.1356)
$PROD_{t-1}^s$	0.4148* (0.2157)	0.1296 (0.2378)	0.2053 (0.1924)
$Age_{65-84,t-1}^s$	1.2532**(0.6252)	0.7730 (0.5487)	0.8150 (0.5187)
$Age_{85+,t-1}^s$	0.0602 (0.1548)	0.2351* (0.1424)	0.2423* (0.1329)
$HSG_{65+,t-1}^s$	-0.0196 (0.0977)	-0.2796**(0.1063)	-0.2894**(0.1032)
c_1	-0.0049 (0.0073)	-0.0076 (0.0069)	-0.0098 (0.0062)
$HSG_{65+,t-2}^s$		-0.0905 (0.1003)	
$HSG_{65+,t-3}^s$		0.0239 (0.0816)	
$HSG_{65+,t-4}^s$		0.0758 (0.0819)	
$HSG_{65+,t-5}^s$		-0.1466* (0.0787)	-0.1378* (0.0707)
R^2	0.6753	0.8164	0.7994
AIC	-53.579	-53.993	-54.125
BIC	-49.621	-48.336	-49.681

Notes: standard errors are in the parentheses

may reflect the cost increasing effect of technology progress on healthcare spending. Higher level of productivity indicates the possibility of providing more advanced and more expensive medical techniques.

The two (stationarized) demographic variables Age^s_{65-84} and Age^s_{85+} have a positive effect, both in the short and medium term. When controlling for higher orders of the lagged elderly's health status, the short term effect of Age^s_{85+} becomes significant at the 90% significance level. This suggests that the increase in the number of the elderly, especially the oldest old, means a larger demand for healthcare. Moreover, we find that the (stationarized) elderly's health status has a clear negative effect. Especially when the higher lagged health status is included, the short term negative effect of the first lagged health status is significant at the 95% confidence level. This suggests that an improved health status of the elderly slows down the growth of per capita healthcare expenditure, indicating that the elderly's health status is an effective indicator of the demand for healthcare.

A Forecasting Study—We forecast health expenditure based on the dynamic relationship between health expenditure and the other eight variables modeled by the VAR framework. We then compare the forecasts of healthcare spending per capita from the three VAR models. Once the VAR has been estimated, we can forecast the future endogenous variables $\boldsymbol{Y}$ for h periods ahead, conditional on time t. For example, h year horizon point forecasts from a VAR(1) model can be written as ($h \geq 1$)

$$\hat{\boldsymbol{Y}}_{t+h} = (\sum_{i=1}^{h-1} \hat{\Theta}_1^i)\boldsymbol{C} + \hat{\Theta}_1^h \boldsymbol{Y}_t \tag{5.7}$$

The VAR model with exogenous variables (VARX, see (5.6)) also can be forecasted when one can predict future paths of the exogenous variables $\boldsymbol{X}$. For example, when two to five year lagged elderly's health status are considered, forecasts of $\hat{\boldsymbol{Y}}_{t+h}$ from a VARX with one lag is

$$\hat{\boldsymbol{Y}}_{t+h} = (\sum_{i=1}^{h-1} \hat{\Theta}_1^i)\boldsymbol{C} + \hat{\Theta}_1^h \boldsymbol{Y}_t + \sum_{s=2}^{5} (\hat{\Phi}_s (\sum_{j=1}^{h} \hat{\Theta}_1^{j-1} HSG^s_{65+,t-s+j})). \tag{5.8}$$

To test the forecasting accuracy, we perform an out-of-sample analysis. This means that the total sample period is cut into a fitting period and a forecasting period. We choose the fitting period from 1972 to 2005, and predict 5 years ahead until the last available sample year 2010. In this way, forecasts of health expenditure in level (PHE_t) can be compared with the realized values from 2006 to 2010. The forecasting accuracy is measured by using the mean square forecasting error (MSFE), in terms of PHE_t, which is the mean square differences between the realized values of PHE_t and the corresponding forecasts. We further compare the VAR(X) forecasts with using a simple

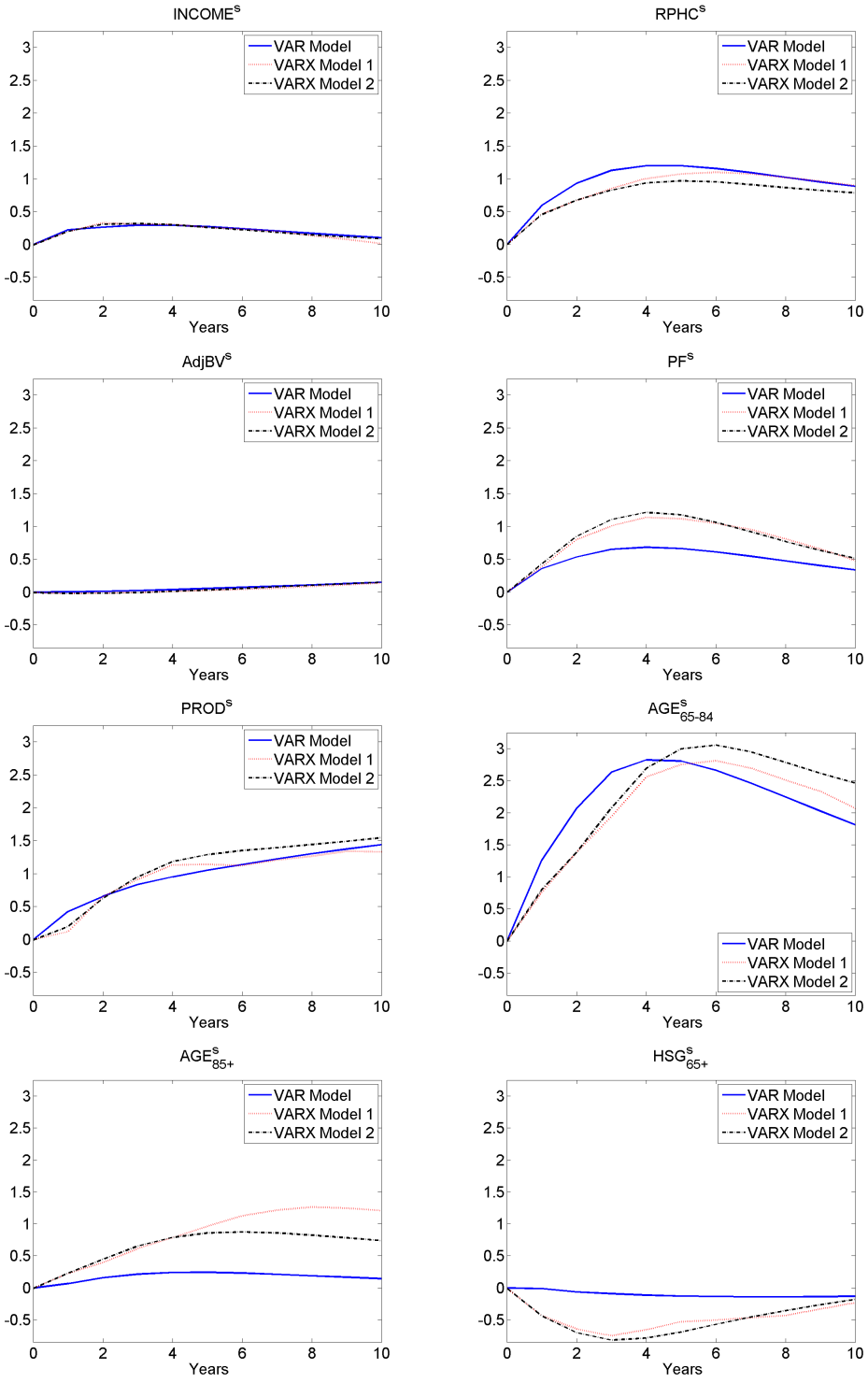

Figure 5.2 – Cumulative Impulse Responses.

This figure shows the cumulative impulse responses on $\log(PHE_t)$ for each of the other eight variables. The main text contains the details.

linear regression. We regress PHE_t^s on the other eight stationarized variables at time t. One difficulty to generate forecasts in this simple linear regression is that it requires to generate first forecasts of the independent variables. Since we focus on investigating the forecasting performance of the VAR(X) for the future healthcare cost, we use the "best" forecasts for the independent variables in the linear regression by using their observed values from 2006 to 2010. In addition, to further investigate the role of the elderly's health status in forecasting the future health expenditure, we also compare our VAR(X) forecasts with forecasts when using a VAR model without including the elderly's health status.

Figure 5.3 presents the in-sample and out-of-sample forecasts for the health expenditure per capita. The in-sample MSFE and out-of-sample MSFE are shown in Table 5.2. The results first suggest that all considered models provide reasonable in-sample forecasts. The out-of-sample foreasts include the period of the financial crisis, starting in 2007. All models have difficulty in forecasting the development of the healthcare spending during this period, including the linear regression model with observed variables. Forecasts based on this appproach seem to overreact to changes in the independent variables. Nevertheless, the VAR model with the higher lagged elderly's health status as observed variables outperforms the other models, both in-sample and out-of-sample. Moreover, including the lagged health status of the elderly provides better in-sample and out-of-sample forecasts than ignoring the previous health status. It is also superior to using a simple linear regression, even when the independent variables take their actual values in the out-of-sample forecasts.

Furthermore, the healthcare spending forecasts based on the VAR system also outperform forecasts provided by the Center for Medicare and Medicaid Services (CMS). We compare our forecasts with forecasts reported in the "National Health Expenditures Projections: 2006 - 2016"[17] and "National Health Expenditures Projections: 2007 - 2017".[18] These reports present the forecasted nominal health expenditure per capita, calculated by CMS as the forecasted real health expenditure times the forecasted price deflator. Using the forecasted nominal health expenditure per capita and the forecasted price deflator from both reports, we find as forecasted real health expenditure per capita in 2010 \$7,477.5 (according to the 2006 - 2016 projections) and \$7,480.2 (according to the 2007 - 2017 projections), using in both cases 2005 as base year. Our predictions are \$7,416.5, \$7,379.6, and \$7,400 based on the three considered VAR systems. These are closer to the actual spending \$7,302.7 in 2010. As most studies on health expenditure projections do, the CMS actuaries assume that healthcare needs are constant across age groups. This assumption means that medical needs change only

[17] Data obtained via private communication, available on request. Poisal, Truffer, Smith, Sisko, Cowan, Keehan, Dickensheets, and the National Health Expenditure Accounts Projections Team (2007) report part of the health expenditure projections based on the "National Health Expenditures Projections: 2006 - 2016".

[18] See Table 1 from http://amcp.org/WorkArea/DownloadAsset.aspx?id=12771

as the age and sex distributions of the population change (Cutler and Sheiner (1998)). One shortcoming of this assumption is that the health condition, which ultimately affects the demand for healthcare, is assumed to be the same as age changes, which is unlikely to be true. Our model relaxes this assumption, by both including the elderly's age structure and the (lagged) elderly's health status, yielding a possible explanation why we obtain a smaller bias in our forecasts.

We conclude that the out-of-sample forecasts based on the dynamic relationship between the variables of interest outperform forecasts without taking into account this relationship, though the in-sample forecasts from the different models are quite similar. Moreover, when taking into account the higher order of the lagged elderly's health status, less biased forecasts are obtained than when ignoring them.

Table 5.2 – Comparison of the mean square forecasting errors among models with different variables

	VAR without HSG	VAR	VARX Model 1	VARX Model 2	Linear Regression
In-sample fit	0.0010	0.0010	0.0007	0.0007	0.0015
Out-of-sample test	0.0051	0.0049	0.0016	0.0028	0.0030

Note: VARX Model 1 is the VAR model including in $\boldsymbol{X}_t$ the exogenous variables $HSG^s_{65+,t-2}$, $HSG^s_{65+,t-3}$, $HSG^s_{65+,t-4}$, and $HSG^s_{65+,t-5}$. VARX Model 2 is the VAR model including the exogenous variables $\boldsymbol{X}_t = HSG^s_{65+,t-5}$.

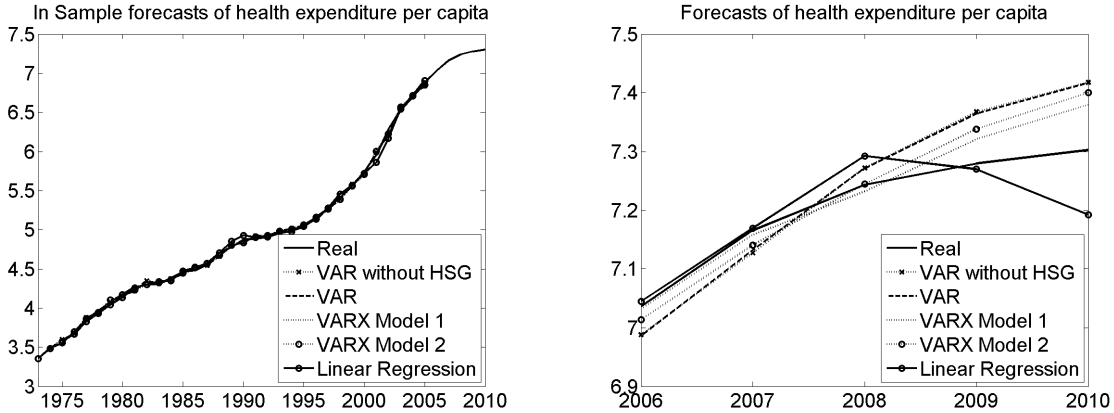

Figure 5.3 – In-sample and out-of-sample forecasts of the per capita healthcare spending

5.5 Conclusion

In this paper we applied a Vector Auto Regression (VAR) model to quantify the dynamic relationship between health expenditure and its determinants. We include a

rather comprehensive set of determinants for healthcare spending, with the aim to reduce a possible bias due to omitted variables. Before applying the VAR model, we first stationarized the variables, using variable specific stationarizing transformations. In terms of the stationarized variables, we find that healthcare is a necessity good as the income elasticity is lower than one. Furthermore, we find that an increase in the relative price of the healthcare service or more public financing both contribute to rising health expenditure. We separate the elderly's age group into the younger elderly and the oldest old, to take into account that most healthcare expenditure seems to happen during the last years of one's life and the share of the people who are approaching death shifts into the oldest old age group. We find that both elderly groups have positive effects on the per capita healthcare spending. Finally, the improvement in the elderly's health status has a clear negative effect on the per capita healthcare spending both in the short and medium term, especially when we control for the higher order of the lagged health status. Our findings confirm the statement by Stearns and Norton (2004) that health expenditure will be affected by a number of factors, including health status trends. Forecasts of the health spending based on the dynamic relationship are superior to forecasts derived from the Center of Medicare and Medicaid. Forecasts based on a model including the elderly's health status are also better than ignoring it.

Appendix: Transformations to Stationarity

We use the Augmented Dickey Fuller test (ADF) to test the existence of the unit root in the examined time series. In this type of test, Dickey and Fuller (1979) actually consider three model specifications, namely a model without time trend and constant ("AR"), a model with a constant ("ARD"), and a model with both a constant and a time trend ("TS"). These specifications can be written as follows,

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \epsilon_t, \tag{AR}$$

$$\Delta y_t = a + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \epsilon_t, \tag{ADR}$$

$$\Delta y_t = a + bt + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \epsilon_t. \tag{TS}$$

To test whether the time series has a unit root or not, the parameter of interest in all the regression equations is γ. $\gamma = 0$ means the presence of a unit root, and the time series is possibly non-stationary. We can estimate the above equations using OLS, and check the t-statistic for the null hypothesis $H_0 : \gamma = 0$. However, the appropriate critical value compared with the t-statistic is different as the usual critical value, and provided in the Dickey-Fuller tables.

In models "TS" and "ARD", we can further perform the F-test to assess the significance of a joint restriction of the parameters. The null hypothesis in model "TS" is that there is a unit root and there is no time trend ($\gamma = 0$ and $b = 0$). The null in model "ARD" is that there is a unit root and there is no constant ($\gamma = 0$ and $a = 0$). Dickey and Fuller (1981) provide the critical values for these additional F-test. If the F-statistic cannot reject the null hypothesis, then $\{y_t\}$ is possibly $I(1)$ with a drift if the model is "TS", and is possibly $I(1)$ without drift if the model is "ARD".

To select a proper model specification, we follow the Pantula principle and start with the least restrictive model, which includes both a constant and a time trend ("TS"). If a unit root is rejected, the time series can be viewed as stationary under this model specification. Otherwise, we continue to the more restrictive model only with a constant ("ARD"). Again, if a unit root is rejected, we can possibly conclude that the time series is stationary under "ARD". Otherwise, the most restrictive model specification without the constant ("AR") will be further tested. However, "AR" may not be a good choice in practice if a clear trend can be observed in the variable.

An important practical issue for the implementation of the ADF test is the specification of the lag length p. An inappropriate choice of p may bias the test. Ng and Perron (1995) proposed the procedure to start with a relatively long lag length. Each time, by applying the usual t-test and/or F-test for the highest lag order, one can decide

whether to reduce the number of lag by one. Enders (2009) suggest that Ng-Perron procedure results in stable size of the test and minimal test power loss. A detailed procedure is described in chapter 4 of book Enders (2009). First, we set an upper bound p_{max} for p. Next, we estimate the ADF test regression with $p = p_{max}$. If the absolute value of the t-statistic for testing the significance of the last lagged difference is greater than 1.6 then we set $p = p_{max}$ and perform the unit root test. Otherwise, we reduce the lag length by one and repeat the process. Enders (2009) illustrates that in the above ADF model specification we can further perform a test for the null hypothesis $\beta_i = \beta_{i+1} = \ldots = \beta_p = 0$ using the standard F-tests to determine the lag length.

In addition to use the t-test and F-test, it is also possible to determine the lag length following the information criteria, such as Akaike information criterion (Akaike (1974)) and Bayesian information criterion or Schwarz criterion (Schwert (1989)). In this analysis, we follow both the Ng-Perron "general to specific" procedure and information criteria (AIC and BIC) to determine the lag length of each model.

Furthermore, since residuals in an appropriately specified model should not have any strong serial correlation, we shall apply a diagnostic check on the residuals in the proposed model specification and lag length. To do so, we perform the Ljung-Box Q-statistic to check if there is any significant autocorrelation among the residuals.

As shown by Figure 5.1, all variables in levels seems to be trended upward. Therefore, we test the unit root for variables in level starting from the model specification "TS", and continue with "ARD" if trend stationary is rejected. We do not apply "AR" specification when analyzing variables in level, since all variables seems to have clear trends. The testing results for each variable are now discussed in detail. In the subsequent analysis, the maximum lag length of 10 is examined, i.e., $p_{max} = 10$.

The logarithm of health expenditure per capita—When analyzing the $\log(PHE)$ in levels, with the model specification "TS", AIC and BIC select the lag length of 1 ($p = 1$). Whereas following the Ng-Perron procedure, the lag length of 3 ($p = 3$) is suggested. As Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in residuals when $p = 1$ and $p = 3$, we further perform the F-test to check whether the coefficients in front of the lagged first difference variables are jointly significant or not. For this purpose, the equation with the lag length $p = 4$ is estimated. The F-statistics are 1.5585 with a p-value of 0.2138 for the null hypothesis $\beta_3 = \beta_4 = 0$, 1.6924 with a p-value of 0.1808 for $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$, and 12.4182 with a p-value of 0.0000 for $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. This indicates that the F-test also selects a model with 1 lag at the 5% significance level.

Therefore, we perform the ADF test for the series in level with $p = 1$ in model specification "TS". Results are presented by the first panel in Table 5.3. Results show that $\log(PHE)$ in level has a unit root if choosing $p = 1$. It is not trend stationary. The F-statistics in the ADF test cannot reject the null hypothesis that $a = b = \gamma = 0$,

suggesting that $\log(PHE)$ is possibly a $I(1)$ process with drift. Following the Pantula principle, we continue to the less restrictive model "ARD", the ADF test still cannot reject the unit root in time series.

Next, we test for a unit root in the first difference. The ADF test results are presented in the second panel of Table 5.3. We again start from the least restrictive model "TS". AIC and BIC, and the Ng-Perron procedure all suggest lag length of 0 ($p = 0$). The ADF-statistic cannot reject the null hypothesis of a unit root. We continue to model only with a constant ("ARD"). In this case, $p = 0$ is again suggested. The ADF-statistic rejects the null. Therefore, we shall proceed under the assumption that the first difference of $\log(PHE)$ is stationary. That is $\log(PHE)$ is a $I(1)$ process.

Table 5.3 – Unit root test for $\log(PHE)$

	"TS"		"ARD"	
H_0	$b = \gamma = 0$	$\gamma = 0$	$a = \gamma = 0$	$\gamma = 0$
	Level			
$p = 1$	2.308(0.680)	-2.148(0.513)	2.489(0.356)	-0.387(0.901)
	First Difference			
$p = 0$	4.723(0.178)	-3.045(0.135)	5.247(0.041)	-3.118**(0.034)

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of income per capita—For $\log(INCOME)$ in levels, in model "TS", AIC and BIC select the lag length of 0 ($p = 0$), whereas the t-tests suggest a lag length of 10 ($p = 10$). The Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in residuals of both lag choices. We further use F-tests in model "TS" with maximum 11 lags, and test the joint restriction of the parameters in front the lagged first differences, starting with testing the last two lagged coefficients until testing all the lagged coefficients. The F-tests suggest a lag length of 0 ($p = 0$) at the 5% significance level. Therefore, we proceed to test the unit root of $\log(INCOME)$ with $p = 0$. Results are shown in Table 5.4. At the 5% significance level, the null hypothesis of a unit root cannot be rejected.

Next, we test for a unit root in the first difference of $\log(INCOME)$. Under "TS", AIC and BIC suggest lag length of 0 ($p = 0$). The t-tests and F-tests following the Ng-Perron procedure also suggest a lag length 0 ($p = 0$). We find that ADF-statistics reject the null of a unit root. Therefore, $\log(INCOME)$ is assumed to be integrated with order 1.

Table 5.4 – Unit root test for $\log(INCOME)$

	"TS"		"ARD"	
H_0	$b = \gamma = 0$	$\gamma = 0$	$a = \gamma = 0$	$\gamma = 0$
		Level		
$p = 0$	1.504(0.865)	-0.917(0.943)	16.105**(0.001)	-1.576(0.479)
		First Difference		
$p = 0$	14.567**(0.001)	-5.384**(0.001)		

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of relative healthcare price—The AIC and BIC select the lag length of 1 ($p = 1$), and the t-tests suggest a lag length of 10 ($p = 10$) when testing $\log(RPHC)$ in levels with model "TS". F-tests further suggest a lag length of 1 ($p = 1$) at the 5% significance level. The Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in residuals when $p = 1$. Therefore, we proceed to test the unit root of $\log(RPHC)$ with $p = 1$. The results are presented in Table 5.5. At the 5% significance level, the null hypothesis of a unit root cannot be rejected. When checking the model specification "ARD", the ADF test again cannot reject the null hypothesis of a unit root.

Next, we test for a unit root in the first difference of $\log(RPHC)$ with $p = 0$ in the model specification "TS", which is suggested by AIC, BIC, and t-tests and F-tests. We find that ADF-statistics cannot reject the null of a unit root. When continuing with the model specifications "ARD" and "AR", the ADF test statistics still suggest that there exists a unit root. Therefore, we further test its second difference.

In the analysis of $\log(RPHC)$ in its second difference, AIC, BIC, and the t-tests and F-tests all select a lag length of 0 ($p = 0$) in the model specification "TS". ADF-statistics now reject the null hypothesis that there is a unit root. Therefore, we conclude here that $\log(RPHC)$ is a $I(2)$ process.

Table 5.5 – Unit root test for $\log(RPHC)$

	"TS"		"ARD"		"AR"
H_0	$b = \gamma = 0$	$\gamma = 0$	$a = \gamma = 0$	$\gamma = 0$	$\gamma = 0$
			Level		
$p = 1$	3.686(0.381)	-1.533(0.799)	5.996**(0.025)	-2.5563(0.112)	
			First Difference		
$p = 0$	5.645(0.104)	-3.052(0.133)	3.041(0.200)	-2.413(0.146)	-1.115(0.238)
			Second Difference		
$p = 0$	13.493**(0.002)	-5.194**(0.001)			

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

The Adjusted Baumol's Variable—For the adjusted Baumol's variable in levels, in model "TS", the AIC selects the lag length of 7 ($p = 7$), the BIC selects the lag length of 5 ($p = 5$), and the t-tests suggest a lag length of 10 ($p = 10$). Again, we use the F-tests to further determine the lag length together with the t tests. The lag length of 0 ($p = 0$) at 5% significance level is suggested. Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in residuals of all the above lag choices. Therefore, we further proceed to test the unit root of $AdjBV$ for scenarios that $p = 0$, $p = 5$, and $p = 7$. Table 5.6 shows the test results. The lag length does not seem to make a difference: At the 5% significance level, the null hypothesis of a unit root is rejected with all three possible choices of lag length. Therefore, we proceed by assuming that $AdjBV$ is a trend stationary process.

Table 5.6 – Unit root test for $AdjBV$

	"TS"	
H_0	$b = \gamma = 0$	$\gamma = 0$
	Level	
$p = 0$	73.97**(0.001)	-4.797**(0.003)
$p = 5$	22.89**(0.001)	-6.708**(0.001)
$p = 7$	11.18**(0.006)	-4.326**(0.009)

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of ratio of public health expenditure to total health expenditure—For $\log(PF)$ in levels, in model "TS", AIC, BIC, t-tests, and F-tests all suggest a lag length of 1 ($p = 1$). Results are shown in Table 5.7. At the 5% significance level, ADF-statistics cannot reject the null hypothesis of a unit root. Therefore, we further check the model specification "ARD". Now, AIC, and the t-tests and F-tests choose the lag length of

1, whereas BIC chooses the lag length of 0. However, when $p = 0$, the Ljung-Box Q-statistics reject the null that there is no autocorrelation in residuals. Therefore, we choose $p = 1$ in model "ARD" to perform the ADF test. The ADF tests again cannot reject the null hypothesis of a unit root.

Next, we test for a unit root in the first difference of $\log(PF)$. $p = 0$ is selected in model "TS". We find that ADF-statistics reject the null of a unit root. Therefore, $\log(PF)$ is assumed to be a $I(1)$ process.

Table 5.7 – Unit root test for $\log(PF)$

	"TS"			
H_0	$b = \gamma = 0$	$\gamma = 0$	$a = \gamma = 0$	$\gamma = 0$
	Level			
$p = 1$	4.908(0.160)	-3.110(0.119)	1.390(0.668)	-0.872(0.780)
	First Difference			
$p = 0$	6.708(0.056)	-3.636**(0.040)		

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of productivity per capita— For $\log(PROD)$ in levels, in model "TS", AIC and BIC select the lag length of 1 ($p = 1$). The t-tests suggest a lag length of 9 ($p = 9$). The F-tests suggest a lag length of 0 ($p = 0$) at the 5% significance level. Furthermore, the Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in the residuals of all the above lag choices. Therefore, we further proceed to test the unit root of $\log(PROD)$ for the scenarios that $p = 0$ and $p = 1$. Results are shown in Table 5.8. At the 5% significance level, the null hypothesis of a unit root cannot be rejected with both possible lag length.

Next, we test for a unit root in the first difference of $\log(PROD)$. We again start from the least restrictive model "TS". AIC and BIC suggest lag length of 1 ($p = 1$). The t-tests and F-tests suggest a lag length of 0 ($p = 0$). Regardless of the lag length, we find that the ADF-statistics reject the null of a unit root. Therefore, $\log(PROD)$ is assumed to be integrated with order 1.

Table 5.8 – Unit root test for $\log(PROD)$

	"TS"		"ARD"	
H_0	$b = \gamma = 0$	$\gamma = 0$	$a = \gamma = 0$	$\gamma = 0$
	Level			
$p = 0$	4.954(0.155)	-2.098(0.536)	17.191**(0.001)	0.825(0.993)
$p = 1$	8.031**(0.026)	-3.203(0.100)		
$p = 2$			5.747**(0.030)	1.606(0.999)
	First Difference			
$p = 0$	11.087**(0.005)	-4.650**(0.004)		
$p = 1$	8.097**(0.025)	-3.991**(0.018)		

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of ratio of population 65 to 84 years old to total population—For $\log(Age_{65-84})$ in levels, in model "TS", AIC, BIC, and the t-tests and F-tests select the lag length of 1 ($p = 1$) at the 5% significance level. The Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in the residuals. Therefore, we further proceed to test the unit root of $\log(Age_{65-84})$ with $p = 1$. Results are shown in Table 5.9. At the 5% significance level, the null hypothesis of a unit root cannot be rejected.

Next, we test for a unit root in the first differences of $\log(Age_{65-84})$. All lag selection criteria suggest a lag length of 0 ($p = 0$). We find that ADF-statistics cannot reject the null of a unit root with model specification "TS". By further performing the test with specification "ARD" and "AR", we still cannot reject the null of a unit root. Therefore, the first difference of $\log(Age_{65-84})$ is not stationary, we need to continue testing its second difference.

In the analysis of the second differences of $\log(Age_{65-84})$, AIC, BIC, and the t-tests and F-tests all select a lag length of 0 ($p = 0$) in the model specification "TS". The ADF-statistics reject the null hypothesis that there is a unit root. Therefore, we conclude here that $\log(Age_{65-84})$ is a $I(2)$ process.

Table 5.9 – Unit root test for $\log(Age_{65-84})$

	"TS"		"ARD"		"AR"
H_0	$b = \gamma = 0$	$\gamma = 0$	$a = \gamma = 0$	$\gamma = 0$	$\gamma = 0$
			Level		
$p = 1$	6.505(0.063)	-3.484(0.056)	2.809(0.265)	-2.288(0.181)	
			First Difference		
$p = 0$	0.844(0.983)	-0.230(0.990)	0.521(0.942)	-1.021(0.716)	-0.845(0.336)
			Second Difference		
$p = 0$	15.546**(0.001)	-5.573**(0.001)			

Notes: *p*-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of ratio of population 85 years and over to total population—The lag selection criteria suggest a lag length of 2 ($p = 2$) at the 5% significance level for $\log(Age_{85+})$ in levels, in model "TS". The Ljung-Box Q-statistics indicate no autocorrelation in residuals. Test restuls are shown in Table 5.10. At the 5% significance level, the null hypothesis of a unit root is rejected. Therefore, we proceed here under the assumption that $\log(Age_{85+})$ is a trend stationary process.

Table 5.10 – Unit root test for $\log(Age_{85+})$

	"TS"	
H_0	$b = \gamma = 0$	$\gamma = 0$
	Level	
$p = 2$	7.105**(0.045)	-3.664**(0.039)

Notes: *p*-values in parentheses. "**" represents that it is significant at the 5% significant level.

The logarithm of share of the elderly population in good health—Finally, we test $\log(HSG_{65+})$ in levels. Under the model specification "TS", the lag length of 0 ($p = 0$) at the 5% significance level is selected. The Ljung-Box Q-statistics cannot reject the null that there is no autocorrelation in the residuals. At the 5% significance level, the null hypothesis of a unit root is rejected. Therefore, we proceed here under the assumption that $\log(HSG_{65+})$ is a trend stationary process. See the results in Table 5.11.

Table 5.11 – Unit root test for $\log(HSG_{65+})$

	"TS"	
H_0	$b = \gamma = 0$	$\gamma = 0$
	Level	
$p = 0$	10.313**(0.007)	-4.519**(0.005)

Notes: p-values in parentheses. "**" represents that it is significant at the 5% significant level.

Detrending—Finally, we detrend variables which are found to be trend stationary by fitting a linear or quadratic function to the trend and substract the fitted values from each observation. We detrend $AdjBV$ and $\log(Age_{85+})$ with a quadratic trend and $\log(HSG_{65+})$ with a linear trend, and in this way obtain stationary deviations of the trends ν_{AdjBV}, $\nu_{Age_{85+}}$, and $\nu_{HSG_{65+}}$, based on the following estimation results:

$$\begin{aligned} AdjBV_t &= 0.2427 + 1.2251t - 0.0188t^2 + \nu_{AdjBV,t} \\ &\quad (0.4622)\ (0.0533)\ (0.0013) \\ \log(HSG_{65+,t}) &= -4.9503 + 0.0339t - 0.0003t^2 + \nu_{HSG_{65+,t}} \\ &\quad (0.0111)\ (0.0013)\ (0.0000) \\ \log(Age_{85+,t}) &= -0.4011 + 0.0030t + \nu_{Age_{85+,t}} \\ &\quad (0.0059)\ (0.0003) \end{aligned}$$

Table 5.12 – Estimates of VAR(1) model for health expenditure, equation (5.2)

$$Y_t = C + \Theta_1 Y_{t-1} + \nu_t$$

$$\hat{C}' = \begin{pmatrix} -0.004895 & 0.028930 & -0.000840 & -0.177070 & 0.023827 & 0.018783 & -0.002903 & 0.006985 & 0.013644 \\ (0.00733) & (0.01814) & (0.00814) & (0.19556) & (0.01432) & (0.00763) & (0.00216) & (0.00699) & (0.01248) \end{pmatrix}$$

$$\hat{\Theta}_1' = \begin{pmatrix} 0.660199 & -0.313795 & 0.152837 & 13.08201 & -0.232513 & -0.356737 & 0.048923 & -0.044764 & -0.046758 \\ (0.19003) & (0.47027) & (0.21118) & (5.07083) & (0.37136) & (0.19777) & (0.05601) & (0.18115) & (0.32372) \\ 0.217141 & -0.188496 & 0.034702 & 0.770658 & -0.156917 & -0.198893 & 0.035488 & -0.088689 & -0.048297 \\ (0.09120) & (0.22569) & (0.10135) & (2.43356) & (0.17822) & (0.09491) & (0.02688) & (0.08693) & (0.15536) \\ 0.587843 & -0.234577 & 0.027727 & -1.991805 & -0.750844 & 0.152168 & 0.126446 & -0.055542 & -0.054656 \\ (0.19711) & (0.48778) & (0.21904) & (5.25966) & (0.38519) & (0.20514) & (0.05809) & (0.18789) & (0.33578) \\ -0.003414 & -0.005300 & 0.001693 & 1.062508 & -0.004468 & -0.008724 & 0.001154 & 0.003837 & -0.024716 \\ (0.00391) & (0.00966) & (0.00434) & (0.10421) & (0.00763) & (0.00406) & (0.00115) & (0.00372) & (0.00665) \\ 0.348931 & -0.380547 & -0.017972 & 2.068272 & -0.067841 & -0.076596 & 0.064679 & -0.029508 & -0.321983 \\ (0.16426) & (0.40650) & (0.18254) & (4.38323) & (0.32101) & (0.17095) & (0.04841) & (0.15658) & (0.27983) \\ 0.414789 & 0.240214 & -0.236797 & -6.719372 & -0.764777 & 0.496402 & 0.049119 & -0.178950 & -0.801989 \\ (0.21565) & (0.53366) & (0.23965) & (5.75439) & (0.42143) & (0.22443) & (0.06356) & (0.20557) & (0.36736) \\ 1.253237 & -2.518390 & 0.477831 & 1.350199 & -0.457598 & 0.734811 & 0.036267 & 0.834635 & 0.276878 \\ (0.62518) & (1.54713) & (0.69476) & (16.6825) & (1.22175) & (0.65065) & (0.18426) & (0.59595) & (1.06501) \\ 0.060155 & -0.009252 & 0.013883 & 0.383768 & -0.044464 & 0.165453 & -0.055178 & 0.878583 & 0.416956 \\ (0.15478) & (0.38304) & (0.17201) & (4.13026) & (0.30248) & (0.16109) & (0.04562) & (0.14755) & (0.26368) \\ -0.019606 & -0.094539 & 0.069489 & 3.493889 & -0.003018 & -0.074537 & -0.033830 & 0.175900 & 0.020641 \\ (0.09773) & (0.24185) & (0.10860) & (2.60780) & (0.19098) & (0.10171) & (0.02880) & (0.09316) & (0.16648) \end{pmatrix}$$

Table 5.13 – Estimates of VARX Model 1 for health expenditure, equation (5.6)

$$\boldsymbol{Y}_t = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \sum_{i=2}^{5} \Phi_i \nu_{HSG_{65+,t-i}} + \nu_t$$

$$\hat{\boldsymbol{C}}' = \begin{pmatrix} -0.007608 & 0.025446 & -0.008649 & -0.171700 & 0.027060 & 0.022748 & -0.003139 & 0.005029 & 0.011082 \\ (0.00694) & (0.02095) & (0.00857) & (0.21331) & (0.01682) & (0.00904) & (0.00280) & (0.00796) & (0.01496) \end{pmatrix}$$

$$\hat{\Theta}_1' = \begin{pmatrix} 0.957844 & 0.066417 & 0.170556 & 9.461959 & -0.483492 & -0.319742 & 0.063692 & -0.178276 & 0.206237 \\ (0.19025) & (0.57444) & (0.23488) & (5.84800) & (0.46121) & (0.24774) & (0.07667) & (0.21810) & (0.41026) \\ 0.204424 & -0.236434 & 0.049071 & 1.556067 & -0.123777 & -0.227010 & 0.034601 & -0.057010 & -0.076676 \\ (0.07702) & (0.23255) & (0.09508) & (2.36738) & (0.18671) & (0.10029) & (0.03104) & (0.08829) & (0.16608) \\ 0.473729 & -0.603419 & -0.133145 & -2.811392 & -0.446766 & 0.162454 & 0.130422 & -0.021148 & 0.101485 \\ (0.18340) & (0.55374) & (0.22642) & (5.63727) & (0.44459) & (0.23881) & (0.07391) & (0.21024) & (0.39548) \\ -0.009830 & -0.014443 & 0.009583 & 1.232042 & -0.000719 & -0.014840 & 0.000936 & 0.010188 & -0.032112 \\ (0.00438) & (0.01323) & (0.00541) & (0.13468) & (0.01062) & (0.00571) & (0.00177) & (0.00502) & (0.00945) \\ 0.390980 & -0.406778 & 0.015346 & 1.481450 & -0.017708 & -0.140271 & 0.070679 & 0.004342 & -0.210976 \\ (0.15120) & (0.45652) & (0.18666) & (4.64751) & (0.36653) & (0.19688) & (0.06093) & (0.17333) & (0.32604) \\ 0.129561 & -0.121344 & 0.257522 & -1.071144 & -0.619071 & 0.188605 & 0.042308 & 0.139784 & -1.037654 \\ (0.23784) & (0.71815) & (0.29364) & (7.31096) & (0.57659) & (0.30971) & (0.09585) & (0.27266) & (0.51289) \\ 0.773048 & -3.467440 & 0.581743 & 14.97261 & 0.050109 & 0.398230 & 0.024944 & 1.180572 & -0.183494 \\ (0.54869) & (1.65671) & (0.67740) & (16.8658) & (1.33014) & (0.71447) & (0.22113) & (0.62900) & (1.18320) \\ 0.235093 & 0.285689 & -0.018272 & -3.683907 & -0.199189 & 0.244364 & -0.048645 & 0.771682 & 0.595457 \\ (0.14241) & (0.43000) & (0.17582) & (4.37756) & (0.34524) & (0.18544) & (0.05739) & (0.16326) & (0.30710) \\ -0.279641 & -0.433390 & 0.094174 & 4.022300 & 0.236730 & -0.107630 & -0.039201 & 0.276646 & -0.011077 \\ (0.10625) & (0.32081) & (0.13118) & (3.26598) & (0.25758) & (0.13835) & (0.04282) & (0.12180) & (0.22912) \end{pmatrix}$$

$$\hat{\Phi}_2' = \begin{pmatrix} -0.090468 & -0.114379 & 0.181171 & 5.891145 & 0.041057 & -0.228893 & -0.003291 & 0.159597 & -0.227004 \\ (0.10026) & (0.30272) & (0.12378) & (3.08182) & (0.24305) & (0.13055) & (0.04041) & (0.11494) & (0.21620) \end{pmatrix}$$

$$\hat{\Phi}_3' = \begin{pmatrix} 0.023901 & -0.144640 & 0.109242 & 3.098722 & 0.022783 & -0.061515 & -0.002017 & 0.086854 & -0.111305 \\ (0.08155) & (0.24623) & (0.10068) & (2.50669) & (0.19769) & (0.10619) & (0.03287) & (0.09349) & (0.17585) \end{pmatrix}$$

$$\hat{\Phi}_4' = \begin{pmatrix} 0.075845 & 0.087653 & -0.090705 & 2.348374 & -0.149851 & 0.093049 & -0.004284 & -0.090411 & -0.163819 \\ (0.08191) & (0.24732) & (0.10112) & (2.51777) & (0.19857) & (0.10666) & (0.03301) & (0.09390) & (0.17663) \end{pmatrix}$$

$$\hat{\Phi}_5' = \begin{pmatrix} -0.146607 & -0.148182 & -0.031679 & -1.217043 & 0.217033 & -0.014912 & -0.009020 & 0.103944 & -0.032944 \\ (0.07871) & (0.23765) & (0.09717) & (2.41938) & (0.19081) & (0.10249) & (0.03172) & (0.09023) & (0.16973) \end{pmatrix}$$

Table 5.14 – Estimates of VARX Model 2 for health expenditure, equation (5.6)

$$\boldsymbol{Y}_t = \boldsymbol{C} + \Theta_1 \boldsymbol{Y}_{t-1} + \Phi_5 \nu_{HSG_{65+},t-5} + \nu_t$$

$$\hat{C}' = \begin{pmatrix} -0.009819 & 0.019885 & -0.003012 & -0.121856 & 0.031373 & 0.017434 & -0.003101 & 0.010085 & 0.011039 \\ (0.00617) & (0.01806) & (0.00790) & (0.20942) & (0.01453) & (0.00838) & (0.00238) & (0.00728) & (0.01357) \end{pmatrix}$$

$$\hat{\Theta}_1' = \begin{pmatrix} 0.995261 & 0.117546 & 0.127609 & 11.57861 & -0.578275 & -0.283720 & 0.060567 & -0.222532 & 0.075953 \\ (0.17739) & (0.51889) & (0.22701) & (6.01712) & (0.41739) & (0.24070) & (0.06845) & (0.20921) & (0.39002) \\ 0.210834 & -0.209645 & 0.022075 & 0.986667 & -0.135998 & -0.202307 & 0.034805 & -0.080280 & -0.059203 \\ (0.07408) & (0.21668) & (0.09480) & (2.51269) & (0.17430) & (0.10052) & (0.02859) & (0.08736) & (0.16287) \\ 0.466546 & -0.579284 & -0.122046 & 0.262360 & -0.507678 & 0.121479 & 0.127298 & -0.014560 & -0.054894 \\ (0.16843) & (0.49269) & (0.21555) & (5.71333) & (0.39632) & (0.22855) & (0.06500) & (0.19864) & (0.37033) \\ -0.008710 & -0.009875 & 0.004346 & 1.053967 & -0.001483 & -0.009527 & 0.001050 & 0.005774 & -0.025175 \\ (0.00351) & (0.01025) & (0.00449) & (0.11891) & (0.00825) & (0.00476) & (0.00135) & (0.00413) & (0.00771) \\ 0.438757 & -0.317490 & -0.071197 & 2.212564 & -0.121288 & -0.059978 & 0.068387 & -0.076913 & -0.288575 \\ (0.13561) & (0.39669) & (0.17355) & (4.60007) & (0.31910) & (0.18402) & (0.05233) & (0.15994) & (0.29817) \\ 0.205345 & 0.158980 & -0.039580 & -8.231632 & -0.733428 & 0.473680 & 0.045485 & -0.115868 & -0.798301 \\ (0.19241) & (0.56284) & (0.24623) & (6.52676) & (0.45275) & (0.26109) & (0.07425) & (0.22693) & (0.42306) \\ 0.815044 & -3.288896 & 0.352742 & 4.667910 & 0.076958 & 0.647720 & 0.032677 & 0.991593 & 0.251885 \\ (0.51873) & (1.51737) & (0.66383) & (17.5956) & (1.22057) & (0.70388) & (0.20017) & (0.61177) & (1.14053) \\ 0.242323 & 0.256260 & 0.021468 & -0.556581 & -0.237622 & 0.201881 & -0.051532 & 0.800959 & 0.449922 \\ (0.13290) & (0.38875) & (0.17007) & (4.50801) & (0.31271) & (0.18033) & (0.05128) & (0.15674) & (0.29221) \\ -0.289356 & -0.436595 & 0.104703 & 4.312181 & 0.241067 & -0.124546 & -0.039306 & 0.286553 & -0.021613 \\ (0.10319) & (0.30184) & (0.13205) & (3.50021) & (0.24280) & (0.14002) & (0.03982) & (0.12170) & (0.22688) \end{pmatrix}$$

$$\hat{\Phi}_5' = \begin{pmatrix} -0.137849 & -0.172955 & -0.002193 & 1.114365 & 0.187711 & -0.043411 & -0.011196 & 0.125222 & -0.141430 \\ (0.07068) & (0.20676) & (0.09045) & (2.39760) & (0.16632) & (0.09591) & (0.02728) & (0.08336) & (0.15541) \end{pmatrix}$$

BIBLIOGRAPHY

AKAIKE, H. (1974): "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716.

AMIRI, A., AND B. VENTELOU (2012): "Granger causality between total expenditure on health and GDP in OECD: Evidence from the Toda-Yamamoto approach," *Economics Letters*, 116(3), 541–544.

AU, D. W. H., T. F. CROSSLEY, AND M. SCHELLHORN (2005): "The Effect of Health Changes and Long-term Health on the Work Activity of Older Canadians," *Health Economics*, 14, 999–1018.

BAIRD, S., J. FRIEDMAN, AND N. SCHADY (2011): "Aggregate Income Shocks and Infant Mortality in the Developing World," *The Review of Economics and Statistics*, 93(3), 847–856.

BALTAGI, B. H., AND F. MOSCONE (2010): "Health care expenditure and income in the {OECD} reconsidered: Evidence from panel data," *Economic Modelling*, 27(4), 804 – 811, Special Issue on Health Econometrics.

BATES, L. J., AND R. E. SANTERRE (2013): "Does the U.S. health care sector suffer from Baumol's cost disease? Evidence from the 50 states," *Journal of Health Economics*, 32(2), 386–391.

BAUMOL, W. J. (1967): "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *The American Economic Review*, 57(3), 415–426.

BEBBINGTON, A., AND J. SHAPIRO (2006): "Incidence of Poor Health and Long-Term Care: Health Transitions in Europe - Results from the European Community Household Panel Survey and Institutional Data. ENEPRI Research Reports No. 34, 23 December 2006," .

BEHNCKE, S. (2012): "Does retirement trigger ill health?," *Health Economics*, 21(3), 282–300.

BERNDT, E. R., D. M. CUTLER, R. G. FRANK, Z. GRILICHES, J. P. NEWHOUSE, AND J. E. TRIPLETT (2000): "Medical care prices and output," in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, vol. 1 of *Handbook of Health Economics*, chap. 3, pp. 119–180. Elsevier.

BLOMQVIST, A., AND R. A. L. CARTER (1997): "Is health care really a luxury?," *Journal of Health Economics*, 16(2), 207 – 229.

BLOOM, D., D. CANNING, AND G. FINK (2008): "Population Aging and Economic Growth," Pgda working papers, Program on the Global Demography of Aging.

BLOOM, D. E., AND D. CANNING (2005): "Health and Economic Growth: Reconciling the Micro and Macro Evidence," working paper 42, Center on Democracy, Development, and The Rule of Law, Stanford Institute of International Studies.

BLOOM, D. E., D. CANNING, AND G. FINK (2009): "The Graying of Global Population and Its Macroeconomic Consequences," PGDA Working Papers 4709, Program on the Global Demography of Aging.

BLOOM, D. E., D. CANNING, R. K. MANSFIELD, AND M. MOORE (2007): "Demographic change, social security systems, and savings," *Journal of Monetary Economics*, 54(1), 92 – 114, Carnegie-Rochester Conference Series on Public Policy: Economic Consequences of Demographic Change in a Global Economy April 21-22, 2006.

BOUND, J. (1991): "Self-Reported Versus Objective Measures of Health in Retirement Models," *Journal of Human Resources*, 26(1), 106–138.

BOUND, J., M. SCHOENBAUM, T. R. STINEBRICKNER, AND T. WAIDMANN (1999): "The dynamic effects of health on the labor force transitions of older workers," *Labour Economics*, 6(2), 179 – 202.

BOUND, J., AND T. WAIDMANN (2007): "Estimating the Health Effects of Retirements," Working Papers wp168, University of Michigan, Michigan Retirement Research Center.

BOX, G. E. P., AND D. R. COX (1964): "An Analysis of Transformations," *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211–252.

BRANCH, L., A. JETTE, C. EVASHWICK, M. POLANSKY, G. ROWE, AND P. DIEHR (1981): "Toward understanding elders' health service utilization," *Journal of Community Health*, 7(2), 80–92.

BURSTROM, B., AND P. FREDLUND (2001): "Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes?," *Journal of Epidemiology and Community Health*, 55(11), 836–840.

CAIRNS, A. J., D. BLAKE, K. DOWD, G. D. COUGHLAN, D. EPSTEIN, AND M. KHALAF-ALLAH (2011): "Mortality density forecasts: An analysis of six stochastic mortality models," *Insurance: Mathematics and Economics*, 48(3), 355 – 367.

CAIRNS, A. J. G., D. P. BLAKE, K. DOWD, G. COUGHLAN, D. EPSTEIN, A. ONG, AND I. BALEVICH (2007): "A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States," *Heriot-Watt University, and Pensions Institute Discussion Paper*, (PI-0701), Working Paper.

CASE, A., AND C. PAXSON (2005): "Sex Differences in Morbidity and Mortality," *Demography*, 42(2), pp. 189–214.

CATALANO, R. A., AND W. A. SATARIANO (1998): "Unemployment and the likelihood of detecting early-stage breast cancer," *American Journal of Public Health*, 88(4), 586–589.

CAWLEY, J., A. S. MORIYA, AND K. I. SIMON (2011): "The Impact of the Macroeconomy on Health Insurance Coverage: Evidence from the Great Recession," Working Paper 17600, National Bureau of Economic Research.

CHANDRA, A., AND J. S. SKINNER (2011): "Technology Growth and Expenditure Growth in Health Care," NBER Working Papers 16953, National Bureau of Economic Research, Inc.

CHRISTIANSEN, T., M. BECH, AND J. LAURIDSEN (2007): "Demographic Changes and Aggregate Healthcare Expenditure in Europe. ENEPRI Policy Briefs No. 4, 18 December 2007," .

COALE, A., P. DEMENY, AND B. VAUGHAN (1983): *Regional Model Life Tables and Stable Populations*, Second edition. Academic Press, New York.

COE, N., AND M. LINDEBOOM (2008): "Does Retirement Kill You? Evidence from Early Retirement Windows," Discussion Paper 2008-93, Tilburg University, Center for Economic Research.

COE, N. B., AND K. HAVERSTICK (2010): "Measuring the Spillover to Disability Insurance due to the Rise in the Full Retirement Age," Working Paper 2010-21, Boston College.

COE, N. B., AND G. ZAMARRO (2011): "Retirement effects on health in Europe," *Journal of Health Economics*, 30(1), 77–86.

COLOMBIER, C. (2012): "Drivers of health care expenditure: Does Baumol's cost disease loom large?," FiFo Discussion Papers - Finanzwissenschaftliche Diskussionsbeiträge 12-5, University of Cologne, FiFo Institute for Public Economics.

CONTOYANNIS, P., AND A. M. JONES (2004): "Socio-economic status, health and lifestyle," *Journal of Health Economics*, 23(5), 965 – 995.

CONTOYANNIS, P., A. M. JONES, AND N. RICE (2004): "The dynamics of health in the British Household Panel Survey," *Journal of Applied Econometrics*, 19(4), 473–503.

CORRAO, G., V. BAGNARDI, A. ZAMBON, AND C. L. VECCHIA (2004): "A meta-analysis of alcohol consumption and the risk of 15 diseases," *Preventive Medicine*, 38(5), 613 – 619.

COSTA, D. L. (2002): "Changing Chronic Disease Rates and Long-term Declines in Functional Limitation Among Older Men," *Demography*, 64(1), 119–137.

CRIMMINS, E., M. HAYWARD, Y. SAITO, AND N. D. J. KENKYŪJO (1996): *Differentials in Active Life Expectancy in the Older Population of the United States*, no. pp. 111-120 in NUPRI reprint series. Nihon University, Population Research Institute.

CRIMMINS, E. M., M. D. HAYWARD, AND Y. SAITO (1994): "Changing Mortality and Morbidity Rates and the Health Status and Life Expectancy of the Older Population," *Demography*, 31(1), pp. 159–175.

CULYER, A. J. (1989): "Cost Containment in Europe," *Health Care Financing Review*, 10.

CURRIE, I. D., M. DURBAN, AND P. H. C. EILERS (2004): "Smoothing and Forecasting Mortality Rates," *Statistical Modeling*, 4, 279–298.

CUTLER, D., E. MEARA, AND S. RICHARDS-SHUBIK (2011): "Healthy life expectancy: Estimates and implications for Retirement Age Policy," Discussion paper, National Bureau of Economic Research.

CUTLER, D. M., AND L. SHEINER (1998): "Demographics and Medical Care Spending: Standard and Non-Standard Effects," NBER Working Papers 6866, National Bureau of Economic Research, Inc.

DE WAEGENAERE, A., B. MELENBERG, AND Y. YANG (2014): "Linking retirement age to life expectancy effects on healthy life expectancy before and after retirement," *Working Paper*.

DHAVAL, D., I. RASHAD, AND J. SPASOJEVIC (2006): "The Effects of Retirement on Physical and Mental Health Outcomes," Working Paper 12123, National Bureau of Economic Research.

DI MATTEO, L. (2005): "The macro determinants of health expenditure in the United States and Canada: assessing the impact of income, age distribution and time," *Health Policy*, 71(1), 23–42.

DICKEY, D. A., AND W. A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Association*, 74(366), pp. 427–431.

DICKEY, D. A., AND W. A. FULLER (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49(4), 1057–72.

DORMONT, B., M. GRIGNON, AND H. HUBER (2006): "Health expenditure growth: reassessing the threat of ageing," *Health Economics*, 15(9), 947–963.

DOWD, K., A. J. CAIRNS, D. BLAKE, G. D. COUGHLAN, D. EPSTEIN, AND M. KHALAF-ALLAH (2010): "Evaluating the goodness of fit of stochastic mortality models," *Insurance: Mathematics and Economics*, 47(3), 255–265.

DREGER, C., AND H. REIMERS (2005): "Health Care Expenditures in OECD Countries: A Panel Unit Root and Cointegration Analysis," *International Journal of Applied Econometrics and Quantitative Studies*, 2(2), 5–20.

DUGGAN, M., AND S. A. IMBERMAN (2006): "Why Are the Disability Rolls Skyrocketing? The Contribution of Population Characteristics, Economic Conditions, and Program Generosity," in *Health in Older Ages: The Causes and Consequences of Declining Disability among the Elderly*, ed. by D. Cutler, and D. A. Wise. University of Chicago Press, Forthcoming.

DUGGAN, M., P. SINGLETON, AND J. SONG (2007): "Aching to retire? The rise in the full retirement age and its impact on the social security disability rolls," *Journal of Public Economics*, 91(7-8), 1327–1350.

EIBICH, P. (2013): "Stress inducing or relieving? Retirement's causal effect on health," .

ENDERS, W. (2009): *Applied Econometric Times Series*, Wiley Series in Probability and Statistics. Wiley, 3rd edn.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–76.

ERDIL, E., AND I. H. YETKINER (2009): "The Granger-causality between health care expenditure and output: a panel data approach," *Applied Economics*, 41(4), 511–518.

FRYAR, C. D., M. D. CARROLL, AND C. L. OGDEN (2012): "Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults: United States, Trends 1960 - 1962 Through 2009 - 2010," *National Center of Health Statustics, Health E-Stats*, pp. 568–573.

GARCÍA-MUÑOZ, T. M., S. NEUMAN, AND T. NEUMAN (2014): "Subjective Health Status of the Older Population: Is It Related to Country-Specific Economic Development Measures?," Working Papers 2014-02, Department of Economics, Bar-Ilan University.

GERDTHAM, U.-G. (1992): "Pooling international health care expenditure data," *Health Economics*, 1(4), 217–231.

GERDTHAM, U.-G., AND B. JONSSON (2000): "International comparisons of health expenditure: Theory, data and econometric analysis," in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, vol. 1 of *Handbook of Health Economics*, chap. 1, pp. 11–53. Elsevier.

GERDTHAM, U.-G., B. JONSSON, M. MACFARLAN, AND H. OXLEY (1998): "The determinants of health expenditure in the OECD countries: a pooled data analysis.," *Development Health Econ Public Policy*, 6, 113–134.

GERDTHAM, U.-G., AND M. LOTHGREN (2000): "On stationarity and cointegration of international health expenditure and GDP," *Journal of Health Economics*, 19(4), 461–475.

GERDTHAM, U.-G., J. SOGAARD, F. ANDERSSON, AND B. JONSSON (1992): "An econometric analysis of health care expenditure: A cross-section study of the OECD countries," *Journal of Health Economics*, 11(1), 63–84.

GIROSI, F., AND G. KING (2008): *Demographic Forecasting*. Princeton University Press, Princeton.

GLEI, D. A., AND S. HORIUCHI (2007): "The narrowing sex differential in life expectancy in high-income populations: Effects of differences in the age pattern of mortality," *Population Studies*, 61(2), 141–159, PMID: 17558883.

HALL, R. E., AND C. I. JONES (2007): "The Value of Life and the Rise in Health Spending," *The Quarterly Journal of Economics*, 122(1), 39–72.

HANSEN, P., AND A. KING (1996): "The determinants of health care expenditure: A cointegration approach," *Journal of Health Economics*, 15(1), 127–137.

HÁRI, N., A. DE WAEGENAERE, B. MELENBERG, AND T. E. NIJMAN (2008): "Longevity risk in portfolios of pension annuities," *Insurance: Mathematics and Economics*, 42(2), 505–519.

HARTWIG, J. (2008): "What drives health care expenditure?-Baumol's model of 'unbalanced growth' revisited," *Journal of Health Economics*, 27(3), 603 – 623.

——— (2010): "Is health capital formation good for long-term economic growth? - Panel Granger - causality evidence for {OECD} countries," *Journal of Macroeconomics*, 32(1), 314 – 325.

——— (2011): "Can Baumol's Model of Unbalanced Growth Contribute to Explaining the Secular Rise in Health Care Expenditure? An Alternative Test," *Applied Economics*, pp. 173–184.

HARVEY BRENNER, M. (1979): "Mortality and the national economy," *The Lancet*, 314(8142), 568–573.

HAYWARD, M. D., AND W. R. GRADY (1990): "Work and Retirement Among a Cohort of Older Men in the United States, 1966-1983," *Demography*, 27(3), pp. 337–356.

HERNAES, E., S. MARKUSSEN, J. PIGGOTT, AND O. L. VESTAD (2013): "Does retirement age impact mortality?," *Journal of Health Economics*, 32(3), 586 – 598.

HILLEN, T., R. SCHAUB, A. HIESTERMANN, W. KIRSCHNER, AND B.-P. ROBRA (2000): "Self rating of health is associated with stressful life events, social support and residency in East and West Berlin shortly after the fall of the wall," *Journal of Epidemiology and Community Health*, 54(8), 575–580.

IMAI, K., AND S. SONEJI (2007a): "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension," *Journal of the American Statistical Association*, 102, 1199–1211.

——— (2007b): "On the Estimation of Disability-Free Life Expectancy: Sullivan's Method and Its Extension," *Journal of the American Statistical Association*, 102, 1199–1211.

JACOBZONE, S. (2000): "Coping with aging: International challenges," *Health affairs*, 19(3), 213–25, Copyright - Copyright Project Hope May/Jun 2000; Last updated - 2013-10-23.

JAGGER, C., P. BARBERGER-GATEAU, AND J.-M. ROBINE (2005): "Disability in older people - indicators, process and outcomes," *Disability & Rehabilitation*, 27(5), 209 – 212.

JAGGER, C., B. COX, S. LE ROY, AND THE EHEMU TEAM (2006): "Health Expectancy Calculation by the Sullivan Method: A Practical Guide," Discussion Paper 3, EHEMU.

JAGGER, C., R. MATTHEWS, N. SPIERS, C. BRAYNE, A. COMAS-HERRERA, T. ROBINSON, J. LINDESAY, AND P. CROFT (2006): "Compression or expansion of disability? forecasting future disability levels under changing patterns of diseases," Discussion paper, King's Fund.

JONES, A. M., N. RICE, AND J. ROBERTS (2010): "Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS," *Economic Modelling*, 27(4), 866–880.

KING, A., AND P. HANSEN (1996): "Aggregate health care expenditure in the United States: evidence from cointegration tests: a comment," *Applied Economics Letters*, 3(1), 37–39.

LADITKA, S. (1998): "New methods for analyzing active life expectancy," *Journal of Aging and Health*, 10(2), 214–241.

LADITKA, S., AND J. LADITKA (2009): "Active Life Expectancy: A Central Measure of Population Health," in *International Handbook of Population Aging*, ed. by P. Uhlenberg, vol. 1 of *International Handbooks of Population*, pp. 543–565. Springer Netherlands.

LADITKA, S. B., AND J. N. LADITKA (2002): "Recent Perspectives on Active Life Expectancy for Older Women," *Journal of Women & Aging*, 14(1-2), 163–184.

LEE, R. D., AND L. R. CARTER (1992): "Modeling and Forecasting U.S. Mortality," *Journal of the American Statistical Association*, 87, 659–671.

LEE, R. D., AND T. MILLER (2001): "Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality," *Demography*, 38(4), 537–549.

LEE, R. D., AND T. MILLER (2002): "An approach to forecasting health expenditures, with application to the U.S. Medicare System," *Health Services Research*, 37, 1365–1386.

LEU, R. (1986): "The public-private mix and international health care costs," *Public and Private Health Care Services: Complementarities and Conflicts*.

LI, X., AND N. MAESTAS (2008): "Does the Rise in the Full Retirement Age Encourage Disability Benefits Applications? Evidence from the Health and Retirement Study," Working Paper 2008-198, University of Michigan.

LINDEBOOM, M., AND M. KERKHOFS (2009): "Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work," *Journal of Applied Econometrics*, 24(6), 1024–1046.

LIU, Y., A. ARAI, K. KANDA, R. B. LEE, J. GLASSER, AND H. TAMASHIRO (2013): "Gender gaps in life expectancy: generalized trends and negative associations with development indices in OECD countries," *The European Journal of Public Health*, 23(4), 563–568.

LUTKEPOHL, H. (1991): *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.

LÜTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*. Springer Publishing Company, Incorporated.

LUTKEPOHL, H. (2007): *New Introduction to Multiple Time Series Analysis*. Springer, 1st ed. 2006. corr. 2nd printing edn.

MACDONALD, G., AND S. HOPKINS (2002): "Unit root properties of OECD health care expenditure and GDP data," *Health Economics*, 11(4), 371–376.

MACKINNON, J. G., AND L. MAGEE (1990): "Transforming the Dependent Variable in Regression Models," *International Economic Review*, 31(2), 315–39.

MAESTAS, N., AND J. ZISSIMOPOULOS (2010): "How Longer Work Lives Ease the Crunch of Population Aging," *Journal of Economic Perspectives*, 24(1), 139–60.

MAJER, I., R. STEVENS, W. NUSSELDER, J. MACKENBACH, AND P. H. VAN BAAL (2012): "Modeling and Forecasting Health Expectancy: Theoretical Framework and Application," *Demography*, pp. 1–25.

MANTON, K. G., X. GU, AND V. L. LAMB (2006a): "Long-Term Trends in Life Expectancy and Active Life Expectancy in the United States," *Population and Development Review*, 32(1), 81–105.

——— (2006b): "Long-Term Trends in Life Expectancy and Active Life Expectancy in the United States," *Population and Development Review*, 32(1), 81–105.

MANTON, K. G., V. L. LAMB, AND X. GU (2007): "Medicare cost effects of recent U.S. disability trends in the elderly: future implications," *J Aging Health*, 19(3), 359–81.

MCGARRY, K. (2004): "Health and Retirement: Do Changes in Health Affect Retirement Expectations?," *The Journal of Human Resources*, 39(3), 624–648.

MCGEE, D. L., Y. LIAO, G. CAO, AND R. S. COOPER (1999): "Self-reported Health Status and Mortality in a Multiethnic US Cohort," *American Journal of Epidemiology*, 149(1), 41–46.

MCGINNIS, J., AND W. FOEGE (1993): "ACtual causes of death in the united states," *JAMA*, 270(18), 2207–2212.

MICHAUD, P.-C., D. GOLDMAN, D. LAKDAWALLA, Y. ZHENG, AND A. GAILEY (2009): "Understanding the Economic Consequences of Shifting Trends in Population Health," *NBER Working Paper*, (15231).

MOKDAD, A., J. MARKS, D. STROUP, AND J. GERBERDING (2004): "ACtual causes of death in the united states, 2000," *JAMA*, 291(10), 1238–1245.

MOLLA, M. T., D. K. WAGENER, AND J. H. MADANS (2001): "Summary Measures of Population Health: Methods for Calculating Healthy Life Expectancy," *Healthy People 2010 Statistical Notes*, (21), 1–11.

MOSCONE, F., AND E. TOSETTI (2010): "Health expenditure and income in the United States," *Health Economics*, 19(12), 1385–1403.

MUNNELL, A. H., AND J. LIBBY (2007): *Will people be healthy enough to work longer*. Center for Retirement Research at Boston College.

MUNNELL, A. H., K. B. MEME, N. A. JIVAN, AND K. E. CAHILL (2004): "Should we raise Social Security's earliest eligibility age," *Issue in Brief*, 18.

MURTHY, N., AND V. UKPOLO (1994): "Aggregate health care expenditure in the United States: evidence from cointegration tests," *Applied Economics Letters*.

MURTHY, N. R., AND A. OKUNADE (2000): "Managed care, deficit financing, and aggregate health care expenditure in the United States: A cointegration analysis," *Health Care Management Science*, 3(4), 279–285.

MURTHY, V. N. (2012): "A Time-Series Investigation of the U.S. Real Health Expenditure: Evidence from Nonlinear Unit Root Tests," *International Advances in Economic Research*, 18(4), 429–438.

NEUMAN, K. (2008): "Quit Your Job and Get Healthier? The Effect of Retirement on Health," *Journal of Labor Research*, 29(2), 177–201.

NEWHOUSE, J. (1977): "Medical-care expenditure: A cross country survey," *Journal of Human Resources*, 12(1), 115–125.

NEWHOUSE, J. P. (1992): "Medical Care Costs: How Much Welfare Loss?," *The Journal of Economic Perspectives*, 6(3), 3–21.

NG, S., AND P. PERRON (1995): "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90(429), 268–281.

NIH (1998): "NHLBI Obesity Education Initiative. Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report," Report 98-4083, National Institutes of Health.

NIU, G., AND B. MELENBERG (2014): "Trends in Mortality Decrease and Economic Growth," *Demography*, forthcoming.

NUSSELDER, W., C. LOOMAN, H. OYEN, J. ROBINE, AND C. JAGGER (2010): "Gender differences in health of EU10 and EU15 populations: the double burden of EU10 men," *European Journal of Ageing*, 7(4), 219–227.

OECD (2013): "Pensions at a Glance 2013: OECD and G20 Indicators," Discussion paper, OECD Publishing.

OKUNADE, A. (2001): "Cost-Output Relation, Technological Progress, and Clinical Activity Mix of US Hospital Pharmacies," *Journal of Productivity Analysis*, 16(2), 167–193.

OKUNADE, A. A., M. C. KARAKUS, AND C. OKEKE (2004): "Determinants of Health Expenditure Growth of the OECD Countries: Jackknife Resampling Plan Estimates," *Health care management science*, 7(3), 173–83.

OKUNADE, A. A., AND V. N. MURTHY (2002): "Technology as a 'major driver' of health care costs: a cointegration analysis of the Newhouse conjecture," *Journal of Health Economics*, 21(1), 147 – 159.

OLIVERIA MARTINS, J., C. DE LA MAISONNEUVE, AND S. BJORNERUD (2006): "Projecting OECD Health and Long-Term Care Expenditures: What Are the Main Drivers?," OECD Economics Department Working Papers 477, OECD Publishing.

PEDERSEN, P. J., AND P. BINGLEY (2011): "Does early retirement keep you healthy and extend your life?," *European Association of Labour Economists*.

PENG, R., L. LING, AND Q. HE (2010): "Self-rated health status transition and long-term care need, of the oldest Chinese," *Health Policy*, 97(2-3), 259–266.

PITACCO, E., M. DENUIT, S. HABERMAN, AND A. OLIVIERI (2009a): *Modeling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press.

——— (2009b): *Modelling longevity dynamics for pensions and annuity business*. Oxford University Press.

POISAL, J. A., C. TRUFFER, S. SMITH, A. SISKO, C. COWAN, S. KEEHAN, B. DICKENSHEETS, AND THE NATIONAL HEALTH EXPENDITURE ACCOUNTS PROJECTIONS TEAM (2007): "Health Spending Projections Through 2016: Modest Changes Obscure Part D's Impact," *Health Affairs*, 26(2).

PRESTON, S. H., P. HEUVELINE, AND M. GUILLOT (2001): *Demography: Measuring and Modeling Population Processes*. Blackwell Publishers, London.

PRETTNER, K., AND D. CANNING (2013): "Increasing life expectancy and optimal retirement in general equilibrium," *Economic Theory*, pp. 1–27.

QUINN, B. C., R. A. CATALANO, AND E. FELBER (2009): "The Effect of Community-Level Unemployment on Preventive Oral Health Care Utilization," *Health Services Research*, 44(1), 162–181.

REHM, J., G. GMEL, C. SEMPOS, AND M. TREVISAN (2003): "Alcohol-related morbidity and mortality," *Alcohol Research & Health*, 27(1), 39 – 51.

RENSHAW, A. E., AND S. HABERMAN (2006): "A Cohort-based Extension to the Lee-Carter Model for Mortality Reduction Factors," *Insurance: Mathematics and Economics*, 38(3), 557–570.

ROBERTS, J. (1999): "Sensitivity of elasticity estimates for OECD health care spending: analysis of a dynamic heterogeneous data field," *Health Economics*, 8(5), 459–472.

ROBINE, J.-M., I. ROMIEU, AND J.-P. MICHEL (2003): *Trends in Health Expectancies*pp. 75–101. John Wiley & Sons, Ltd.

RUHM, C. J. (2000): "Are Recessions Good For Your Health?," *The Quarterly Journal of Economics*, 115, 617–650.

RUHM, C. J. (2003): "Good times make you sick," *Journal of Health Economics*, 22(4), 637 – 658.

RUHM, C. J. (2004): "Macroeconomic Conditions, Health and Mortality," NBER Working Papers 11007, National Bureau of Economic Research, Inc.

SCHOEN, R., AND K. WOODROW (1980): "Labor force status life tables for the United States, 1972," *Demography*, 17(3), 297–322.

SCHWAN, A., AND E. SAIL (2013): "Assessing the economic and budgetary impact of linking retirement ages and pension benefits to increases in longevity," European Economy. Economic Papers. 512, European Commision.

SCHWERT, G. W. (1989): "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business & Economic Statistics*, 7, 147.

SIMS, C. A. (1980): "Macroeconomics and Reality," *Econometrica*, 48(1), pp. 1–48.

SINGER, B. H., AND K. G. MANTON (1998): "The effects of health changes on projections of health service needs for the elderly population of the United States," *Proceedings of the National Academy of Sciences*, 95, 15618–15622.

SMITH, J. P. (1999): "Healthy Bodies and Thick Wallets: The Dual Relation between Health and Economic Status," *Journal of Economic Perspectives*, 13(2), 145–166.

SOLAKOGLU, E. G., AND A. CIVAN (2012): "Does morbidity matter? Perceived health status in explaining the share of healthcare expenditures," *Applied Economics*, 44(16), 2027–2034.

STEARNS, S. C., AND E. C. NORTON (2004): "Time to Include Time to Death? The Future of Health Care Expenditure Predictions," *Health Economics*, 13, 315–327.

STURM, R. (2002): "The Effects Of Obesity, Smoking, And Drinking On Medical Problems And Costs," *Health Affairs*, (2), 245–253.

SULLIVAN, D. F. (1971): "A Single Index of Mortality and Morbidity," *HSMHA Health Reports*, 86, 347–354.

TEFFT, N., AND A. KAGELEIRY (2014): "State-Level Unemployment and the Utilization of Preventive Medical Services," *Health Services Research*, 49(1), 186–205.

THORSLUND, M., J. WASTESSON, N. AGAHI, M. LAGERGREN, AND M. PARKER (2013): "The rise and fall of womenæŠ¯ advantage: a comparison of national trends in life expectancy at age 65çˆ•ears," *European Journal of Ageing*, pp. 1–7.

TOFFOLUTTI, V., AND M. SUHRCKE (2014): "Assessing the short term health impact of the Great Recession in the European Union: A cross-country panel analysis," *Preventive Medicine*, 64(0), 54 – 62.

TROVATO, F., AND N. B. HEYEN (2006): "A varied pattern of change of the sex differential in survival in the G7 countries," *Journal of Biosocial Science*, 38, 391–401.

TULJAPURKAR, S., N. LI, AND C. BOE (2000): "A Universal Pattern of Mortality Decline in the G7 Countries," *Nature*, 405, 789–792.

UNGER, R., AND A. SCHULZE (2013): "Can We Really (All) Work Longer? Trends in Healthy Life Expectancy According to Social Stratum in Germany," *Comparative Population Studies - Zeitschrift fÃ¼r BevÃ¶lkerungswissenschaft*, 38(3).

VAN BAAL, P., F. PETERS, J. MACKENBACH, AND W. NUSSELDER (2013): "Forecasting educational differences in (disability free) life expectancy," Working Paper.

VAN BAAL, P., F. PETERS, J. MACKENBACH, AND W. NUSSELDER (2013): "Forecasting educational differences in (disability free) life expectancy," Working Paper.

VAN DE WATER, H. P., R. J. PERENBOOM, AND H. C. BOSHUIZEN (1996): "Policy relevance of the health expectancy indicator; an inventory in European Union countries," *Health Policy*, 36(2), 117 – 129.

VAN DER HEIDE, I., R. M. VAN RIJN, S. J. ROBROEK, A. BURDORF, AND K. I. PROPER (2013): "Is retirement good for your health? A systematic review of longitudinal studies," *BMC public health*, 13(1), 1180.

VAN ELK, R., E. MOT, AND P. FRANSES (2009): "Modelling health care expenditures; overview of the literature and evidence from a panel time series model," CPB Discussion Paper 121, CPB Netherlands Bureau for Economic Policy Analysis.

VAN OYEN, H., B. COX, C. JAGGER, E. CAMBOIS, W. NUSSELDER, C. GILLES, AND J.-M. ROBINE (2010): "Gender gaps in life expectancy and expected years with activity limitations at age 50 in the European Union: associations with macro-level structural indicators," *European Journal of Ageing*, 7(4), 229–237.

VAN OYEN, H., W. NUSSELDER, C. JAGGER, P. KOLIP, E. CAMBOIS, AND J.-M. ROBINE (2013): "Gender differences in healthy life years within the EU: an exploration of the 'health - survival' paradox," *International Journal of Public Health*, 58(1), 143–155.

WALDRON, I. (1993): "Recent trends in sex mortality ratios for adults in developed countries," *Social Science & Medicine*, 36(4), 451 – 462.

WANG, Z., AND A. J. RETTENMAIER (2007): "A note on cointegration of health expenditures and income," *Health Economics*, 16(6), 559–578.

WEIL, D. N. (2007): "Accounting for The Effect of Health on Economic Growth," *The Quarterly Journal of Economics*, 122(3), 1265–1306.

WERBLOW, A., S. FELDER, AND P. ZWEIFEL (2007): "Population ageing and health care expenditure: a school of 'red herrings'?," *Health Economics*, 16(10), 1109–1126.

WHO (2011): "Global status report on alcohol and health," Report, World Health Organization.

WILMOTH, J. R. (1993): "Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change," *U.C. Berkeley Technical Report*.

XU, K., P. SAKSENA, AND A. HOLLY (2011): "The determinants of health expenditure," working paper, World Health Organization.

YANG, Y., A. DE WAEGENAERE, AND B. MELENBERG (2013a): "Do Americans Live Longer and Healthier? Forecasting Healthy Life Expectancy by Including Dynamic Evolutions of Mortality, Health, and Macroeconomic Variables," *Working Paper*.

——— (2013b): "Stochastic Modelling and Forecasting of Health Changes in the U.S. Population," *Working Paper*.

YANG, Y., AND B. MELENBERG (2014): "An analysis of the interatiction between health expenditure and its determinants in the U.S.," *Working Paper*.

YANG, Z., E. C. NORTON, AND S. C. STEARNS (2003): "Longevity and Health Care Expenditures: The Real Reasons Older People Spend More," *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 58(1), S2–S10.

ZWEIFEL, P. (1984): "Technological change in health care: Why are opinions so divided?," *Managerial and Decision Economics*, 5(3), 177–182.

ZWEIFEL, P., S. FELDER, AND M. MEIERS (1999): "Ageing of population and health care expenditure: a red herring?," *Health Economics*, 8(6), 485–496.

ZWEIFEL, P., S. FELDER, AND A. WERBLOW (2004): "Population Ageing and Health Care Expenditure: New Evidence on the 'Red Herring'," *The Geneva Papers on Risk and Insurance*, 29(4), 652–666.